TEAM tAO

7 January 2005

Science

Vol. 307 No. 5706
Pages 1–164 $10

DISKS IN SPACE

125 YEARS OF GLOBAL Science

AAAS

Science

Special Issue

Disks in Space

Volume 307
7 January 2005
Number 5706

Artist's view of a collision in the Kuiper belt. The Sun is 45 astronomical units away and shines within the zodiacal dust cloud. Jupiter (pink dot) and Neptune (blue dot) are to the right of the Sun. The Kuiper belt is a remnant of the circumstellar disk that formed the solar system and is a source of comets. [Image: Daniel D. Durda]

30

45

Contents continued

Science

Perspectives

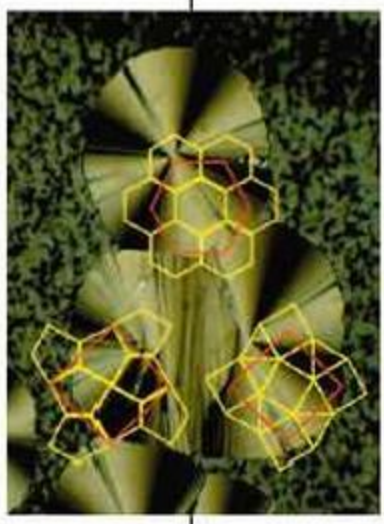
Review

Science Express www.sciencexpress.org

Brevia

Research Article

Reports



96

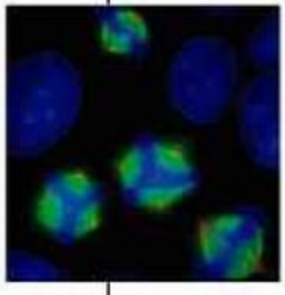
Contents continued

Science

REPORTS CONTINUED

25 & 102

111



127

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals Mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2005 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership and subscription (51 issues): $135 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $550. Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $80. First class, airmail, student, and retired rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069053A. Printed in the U.S.A.

Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. Postmaster: Send change of address to Science, P.O. Box 1811, Danbury, CT 06813–1811. Single copy sales: $10.00 per issue prepaid includes surface postage; bulk rates on request. Authorization to photocopy material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $15.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for Science is 0036-8075/03 $15.00. Science is indexed in the Reader's Guide to Periodical Literature and in several specialized indexes.

Contents continued

sciencenow www.sciencenow.org Daily News Coverage

A Chiral Wildcard
Finding raises questions about long-term effects of common insecticides.

Wake Me When It's Over
A shift from sleeping late to rising early may signal end of adolescence.

Is Allergy a Gut Reaction?
Yeast infection triggered by antibiotics causes an allergic response.

Microbiologist Kristine Brenneman.

science's next wave www.nextwave.org Career Resources for Young Scientists

MiSciNet: Overcoming Odds *E. Francisco*
An environmental microbiologist at Humboldt State University encourages Native Americans and women to enter science.

Canada: From France with a Fellowship *A. Fazekas*
A French postdoc now working at an Agriculture Canada Microbiology laboratory shares her story.

UK: Why Is a Biochemist Teaching Business? *R. Philips*
A former biochemist describes how he became interested in business while still in the lab.

Postdoc Network: California Comes Through! *B. Benderly*
University of California launches its unified health plan for the 6000+ postdocs in the 10-campus system.

GrantsNet: January 2005 Funding News *Edited by S. Martin and S. Otto*
This update features new grant, fellowship, and scholarship opportunities from private and public sources.

Global: *Science*'s Next Wave's Next Wave *J. Austin*
Our new editor describes our audacious agenda for 2005 and beyond.

science's sage ke www.sageke.org Science of Aging Knowledge Environment

Perspective: Report on the 14th Annual Meeting of the German Society for Geriatric Research *C. Scheckhuber*
German scientists present eclectic research on aging.

News Focus: Mix and Mismatch *M. Leslie*
Cancer-promoting protein introduces mistakes in DNA pairing.

News Focus: New Age *R. J. Davenport*
Novel method clocks yeast longevity.

Dissecting mold age.

science's stke www.stke.org Signal Transduction Knowledge Environment

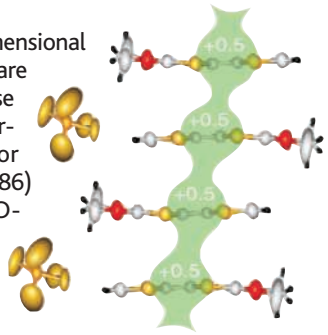
Editorial Guide: 2004 Signaling Breakthroughs of the Year *E. M. Adler, N. R. Gough, L. B. Ray*
Signaling experts nominate notable advances in basic and applied cell signaling research.

Review: The Yins of T Cell Activation *J. O. Liu*
Negative regulators maintain unstimulated T cells in the quiescent state and act as feedback inhibitors to limit T cell–receptor signaling.

Protocol: A Combined Approach for the Localization and Tandem Affinity Purification of Protein Complexes from Metazoans *I. M. Cheeseman and A. Desai*
Proteins can be labeled for both visualization and purification.

Separate individual or institutional subscriptions to these products may be required for full-text access.

Fast Light Switch



Certain organic salts form one-dimensional (1D) or 2D electronic bands that are partially filled and that can give rise to electronic and magnetic properties such as superconductivity or ferroelectricity. **Chollet *et al.*** (p. 86) examined the organic salt $(EDO\text{-}TTF)_2PF_6$, where EDO-TTF is ethylenedioxytetrathiafulvalene, which forms a quasi-1D band that is one-quarter–filled with hole carriers. This material displays a metal-to-insulator (M-I) transition near room temperature that arises from structural changes that lead to charge ordering. The authors find that this M-I transition can be brought about very rapidly (in a few picoseconds) after photoexcitating only a very small fraction of the molecules within the crystal (about 1 in 500) at temperatures near ambient. This phase transition appears to be driven by a coherent phonon generation process caused by the interaction between the electrons and the lattice. Such properties may prove useful as an ultrafast molecular switch.

Seek, Fortify, Then Destroy

In clinical trials, "anti-angiogenic" drugs, which are designed to destroy the blood vessels that feed tumors, have limited efficacy when administered as single agents. However, when provided as a combination therapy, they enhance the efficacy of conventional cytotoxic drugs targeting tumor cells, even though the destruction of the tumor vasculature might be expected to impede drug delivery to the tumor. **Jain** (p. 58) reviews evidence supporting the counterintuitive notion that anti-angiogenic drugs initially fortify, rather than destroy, the tumor vasculature, thereby improving delivery of cytotoxic drugs to the tumor. If further substantiated, this hypothesis would have important implications for the optimal dose and scheduling of combination cancer therapies.

Dissecting Malaria's Genetic Strategies

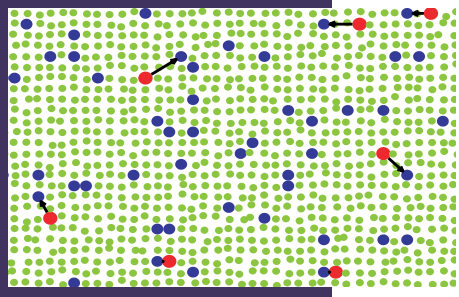
Plasmodium parasites, the agents responsible for malaria, are of intense interest, but they have complex life cycles within their mosquito vectors and within their mammalian hosts that make molecular analysis difficult to untangle. A comparative genome analysis by **Hall *et al.*** (p. 82) shows that, apart from conserved central sections of chromosomes, there are genes evolving rapidly in response to life-cycle, stage-specific pressures. For example, transcriptional profiling and proteomic analysis of several species of parasite has helped tease apart aspects of the little understood sexual cycle of these parasites.

Pentagonal Columnists

Some shapes, like triangles and squares, can regularly pattern or tile a flat space, whereas regular pentagons are only able to tile a sphere. **Chen *et al.*** (p. 96) have synthesized molecules with three incompatible segments that form liquid-crystalline columnar phases. The columns then tile into either identical pentagonal cylinders or a structure composed of square shapes and triangular columns. This packing is possible because of the combination of order and mobility in the fluid state of this type of matter.

Through a Glass Slowly

Electrons moving from donor to acceptor sites often must tunnel through the potential barrier set up by the intervening medium, such as the peptide chains in proteins. Most model studies of these processes have focused on systems in which the donor and acceptor sites are connected by a covalent bridge. **Wenger *et al.*** (p. 99) have explored the effect of nonbonded contacts on tunneling by examining electron transfer rates for random arrays of donors and acceptors in frozen glasses of toluene and 2-methyltetrahydrofuran. The transfer rates are much slower than for covalently bonded alkane bridges at comparable distances.



Salt Survivors

Immense salt deposits beneath the Mediterranean floor are the legacy of its having evaporated to dryness about 6 million years ago. **Van der Wielen *et al.*** (p. 121) have explored the microbiology of deep hypersaline anoxic remnants. A picture emerges of whole microbial communities that are far from being biogeochemical dead-ends. Rather they are contributing to global cycles while thriving in some of the most saline environments known.

Tracing Temple Timing

Several of the Hawaiian islands contain relic temples that were built by their rulers and functioned as centers of control. Radiocarbon dates on wood and charcoal associated with the temples implied that they were built during a 250-year period as the Hawaiian societies evolved and grew. Coral was placed and enclosed in special compartments on these temples as part of their dedication, and **Kirch and Sharp** (p. 102; see the news story by **Stokstad**) dated preserved corals from temples in Maui and Molokai using the ^{230}Th method, which provides more accurate dates for this time. The dates of the coral branches span about 30 years on Maui (just after A.D. 1600) and are slightly older on Molokai. The temples were all completed, and presumably rule was consolidated, much more rapidly than had been believed, perhaps within a single generation.

Bioremediation Bug Genome Revealed

Dehalococcoides ethenogenes is the only bacterium known to reductively dechlorinate groundwater pollutants, tetrachloroethene (PCE) and trichloroethene (TCE), to ethylene. **Seshadri *et al.*** (p. 105) now present an analysis of the genome of *D. ethenogenes*. Multiple dehalogenases and reductases were identified which indicate that the organism is highly evolved to utilize halogenated organic compounds and H_2. The analysis provides insight into the organism's complex nutrient requirements, and surprisingly suggests that an ancestor was a nitrogen-fixing

CONTINUED ON PAGE 15

CREDITS: (TOP TO BOTTOM) CHOLLET *ET AL.*; WENGER *ET AL.*

Published by AAAS

autotroph. Because the organism is difficult to culture, the genome sequence contributes significantly to our understanding of the physiology of this organism and its bioremediation potential.

Picky Eaters

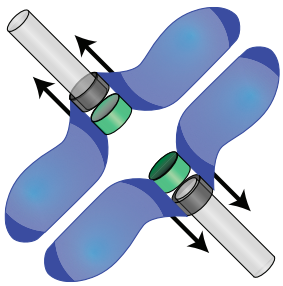
It is widely assumed in foraging theory that predators cannot balance their nutrient intake, but instead maximize their energy intake subject to prey size, abundance, and time constraints. **Mayntz *et al.*** (p. 111) show that this is not the case, using three species of invertebrates (ground beetles, wolf spiders, and web spiders) with widely different feeding biology. When the diet of the predators was manipulated to render them either protein- or lipid-deficient, the animals adjusted their feeding to make good the specific deficit. Compensatory nutrient selection occurred either by selecting among foods of different nutritional composition, by adjusting consumption of a single prey type, or by extracting nutrients selectively from within individual prey items.

Calcium Channels in T Lymphocytes

Calcium represents a critical signaling mediator in a number of biological systems, including excitable cells of such as neurons and in lymphocytes of the immune system. However, the identity of channels that mediate calcium entry in lymphocytes has been unclear. **Badou *et al.*** (p. 177; see Perspective by **Winslow and Crabtree**) find that T cells express two forms of voltage-gated calcium channel (Cav) that are required for mediating activation signals critical for normal T cell functions. Cav activity was increased directly by T cell receptor stimulation.

A Spindle Here, a Spindle There

During cell division, replicated chromosomes align on the mitotic spindle poised to segregate to opposite ends of the cell. To prevent errors during mitosis, a spindle checkpoint monitors proper attachment of chromosomes to the spindle microtubules as well as tension that presumably exists between the chromosomes and the spindle. **Indjeian *et al.*** (p. 130) now describe Sgo1, a protein found on kinetochores (the central region of chromosomes that become attached to the mitotic spindle) that also has a microtubule-binding domain. In mutant yeast lacking Sgo1, chromosomes no longer align correctly on the spindle, and cell cycle progression is blocked. Sgo1 is likely to represent part of the cell's tension sensing machinery when errors in chromosome-spindle interaction occur. Many tumor cells are characterized by increased genomic instability and chromosome segregation defects, and may possess extra microtubule-organizing centrosomes and multipolar mitotic spindles. **Quintyne *et al.*** (p. 127) now find that cytoplasmic dynein-mediated centrosome clustering can help to prevent the formation of multipolar spindles in cells containing additional centrosomes. The authors suggest that the generation of spindle multipolarity in transformation may require two distinct steps—centrosomal amplification followed by centrosome separation.



CREDIT: INDJEIAN *ET AL.*

How Electrons Sink or Swim

Hydrated electrons, which are of importance in radiolytic chemistry and biologically relevant electron transfer, have been studied by using gas-phase water clusters as proxies for bulk water. Do clusters of roughly 50 or more water molecules truly mimic the solvating cavity in the bulk, or do the excess electrons bind to the cluster surface? **Verlet *et al.*** (p. 93, published online 16 December 2004) used photoelectron imaging to garner evidence for two distinct water cluster types, which they assign to structures with either a surface-bound or internally solvated electron. The traditional method of cluster preparation yields the internally solvated structure and supports the applicability of prior studies to the bulk. In contrast, the surface-bound class, with significantly smaller electron binding energies, results from electron attachment to vibrationally colder neutral clusters.

Published by AAAS

A New Year and Anniversary

Another New Year has arrived, and for *Science*, which celebrates the 125th year of its publication, it's a happy anniversary. Please don't worry—we don't plan to salute the occasion with a summary of all the new knowledge that has been introduced in our pages over the past century and a quarter. But we do invite readers to consult the very first issue to get a sense of how much has happened over that time: Volume 1, Number 1, published in July of 1880 (www.sciencemag.org/sciext/firstissue). In fact, every past issue of *Science* can be found by consulting JSTOR, an archive available to any member of the American Association for the Advancement of Science (through the aaasmember.sciencemag.org gateway) or to anyone in a JSTOR-participating institution. The pioneer issues contain some interesting items, including an essay by Thomas Huxley, one of Charles Darwin's admiring scientific colleagues, in which he argues that Darwin's theory of evolution is here to stay. Good call.

What shall we do by way of celebration? We may have a party, but we have a more serious purpose in mind. One of the things that has changed most dramatically since Volume 1, Number 1 is the increasingly significant role played by scientists in countries that, when Huxley was writing about Darwin, were not among the nations in which new experimental work was being done. Today, researchers in the developing world are addressing some of the most interesting and daunting scientific challenges of our time, often under limitations that are not shared by their colleagues in wealthier countries.


SCIENCE

A WEEKLY RECORD OF SCIENTIFIC PROGRESS.

ARTISTS' MATERIALS.

PHOTOGRAPHIC SPECIALTY!

MANUFACTURING OPTICIANS.

MAGIC LANTERNS

STEREOPTICONS

CATALOGUES ON APPLICATION.

We have invited a dozen of the best of these to provide an account of their work, and one of these essays will appear in each month of this anniversary year. We asked them to talk about how they practice their own kind of science rather than about the special scientific needs of their own nations or regions. The first of these, by the South African botanist Patricia Berjak, appears on p. 47 of this issue. It demonstrates clearly the connections between basic research (in this case on seed biology and storage regimes) and the needs of regional ecosystems.

Anniversaries are also a time to look for ways to get better. From time to time, we ask readers of *Science* how they use the journal, what they turn to first, and what difficulties they have with our material. It may not surprise you that if you are typical, you turn first to a Brevia, Report, or Research Article in your own subdiscipline. The next stop is likely to be News, Perspectives, or Policy Forums. After that, perhaps something of interest in Books, or Letters, or even (hopeful thought here) the Editorial page. The discouraging aspect of what we learn is how difficult you find it to access and appreciate original research in areas outside your own. And that's not your fault.

The problem is not unique to *Science*. It is hard for authors to avoid aiming reports of original research at the cognoscenti, especially in fields where movement at the frontier is active. Because the methodological grain of each discipline has become extremely fine, it requires heavy use of technical language, jargon, and acronyms. That tends to make even the title of the average communication in molecular biology in any top-tier journal impenetrable by an ecologist, let alone a physicist. But it's not quite fair to lay the entire problem on complexity, which after all is part of the real world and therefore something we have to deal with. So what can be done?

My colleagues and I had an initial experience with this translation challenge when we began, 2 years ago, to write those one-sentence descriptions of the main result of each paper for inclusion in the Table of Contents. We are still surprised from time to time at how hard it is to communicate the essence of a finding in nontechnical language that can be understood by the nonspecialist. This Week in Science, News, Perspectives, and Editors' Choice are all helpful for providing context and making new research more accessible. We'd like to do more, and Berjak's article in this issue supplies a useful model of how to make a scientific story readable for those outside the specialty.

Donald Kennedy
Editor-in-Chief

10.1126/science.1109092

Published by AAAS

edited by Gilbert Chin

PALEONTOLOGY

Early Toolmakers

A view of the east bank of the Kada Gona River.

The Hadar Formation, exposed by the Awash River in Ethiopia, has yielded hominid fossils spanning several million years, including Lucy (*Australopithecus afarensis*), dated to more than 3 million years ago (Ma). The uppermost part of the Hadar (now designated as the Busidima Formation) also hosts what seem to be the oldest known tools, chiseled river cobbles, and associated debris flakes, dated to about 2.6 Ma. Quade *et al.* document how the environment of the Hadar Formation evolved along with these early hominids. Their analysis shows that the river flowed through forest, mixed with some grassland, which expanded as the climate dried. Early stone tools were collected from cobble bars in the main river and processed nearby, but up on the banks. Later, cobbles were transported farther away. Interestingly, the first occurrence of tools is found above the abrupt appearance of cobbles younger than 3 Ma in the section. These tools may thus represent the appearance of a local resource rather than marking the true technological innovation, which would have happened earlier. — BH

Geol. Soc. Am. Bull. **116**, 1529 (2004).

IMMUNOLOGY

Treating Disease with Worms

Crohn's disease is a debilitating inflammatory condition of the intestine. Although the etiology is unclear, the disease is thought to result from inappropriate activation of the immune system against the bacterial flora of the gut. In developing countries, where infection with parasitic intestinal helminths is widespread, Crohn's disease is rare, leading to the notion that the allergic-like state generated by parasitic worms counteracts proinflammatory influences.

To test this, Summers *et al.* fed Crohn's patients eggs of the common pig helminth *Trichuris suis*, which can colonize the human intestine for short periods without pathology. A marked improvement was seen in most of the patients, and these clinical results are paralleled by the observations of Elliott *et al.*, who found that giving the helminth *Heligmosomoides polygyrus* to mice that were afflicted with a Crohn's-like condition reversed inflammation. In protected animals, there was a redress of the imbalance toward proinflammatory cytokines, and these early results suggest that unconventional therapy of this type might be effective in treating a range of chronic inflammatory diseases that extend beyond the gut. — SJS

Gut **54**, 87 (2005); *Eur. J. Immunol.* **34**, 2690 (2004).

EVOLUTION

Nothing in Common

Analyses of the diversity of marine genera through the Phanerozoic have identified five great global mass extinctions. Bambach *et al.* use Sepkoski's compilation of the stratigraphic ranges of genera at the stage and substage levels to evaluate the continuity of these five big events with background extinction. They see six major temporal intervals of alternating high and low extinction intensity. The Late Devonian and end-Triassic diversity crashes occurred during intervals of generally high extinction and low origination. For these events, extinction intensities—although higher than the average for the inclusive interval—are not distinct outliers, and almost two-thirds of the diversity loss is explained by reduced origination. For the end-Ordovician, end-Permian, and end-Cretaceous events, origination rates exceed those in their temporal neighborhoods, and extinction rates are exceptionally high. These three events appear to differ from each other and from the other two in their physiological selectivity, their ecological impact, and the nature of their effects on particular taxa, and hence are unlikely to be due to a common cause. — SJS

Paleobiology **30**, 522 (2004).

CHEMISTRY

Almost as Bright

Tracking particles and cells in the fluorescence microscope is a key analytical technique in cell biology and materials science. Increasing demand has led to the synthesis and functionalization of new fluorophores and semiconductor nanoparticles (quantum dots). However, many fluorophores are readily photobleached and not very bright, whereas quantum dots require capping layers to prevent aggregation, and their synthesis requires harsh solvents and precursors.

Ow *et al.* have created a hybrid structure with an organic fluorophore covalently attached to a silica precursor, forming an organic core surrounded by a thin silica shell, which is then encapsulated using sol-gel chemistry to make particles 20 to 30 nm in diameter. Adding the outer shell of silica increased the brightness by a factor of 30. One reason is that the shell protects the fluorophore from solvent, which also increases its photostability. The silica nanoparticles are not quite as bright as similarly sized quantum dots, but they can be easily functionalized using the well-established and broad library of silane coupling methods. — MSL

Nano Lett. 10.1021/nl0482478 (2004).

SEISMOLOGY

Urban Hazards

In comparison to the four-fold increase in the world's population to about 6 billion, the percentage of people killed in earthquakes declined only slightly from 1900 to 2000. Although this trend has been assumed to reflect better building codes, Bilham's analysis suggests

CREDITS: SILESHI SEMAW

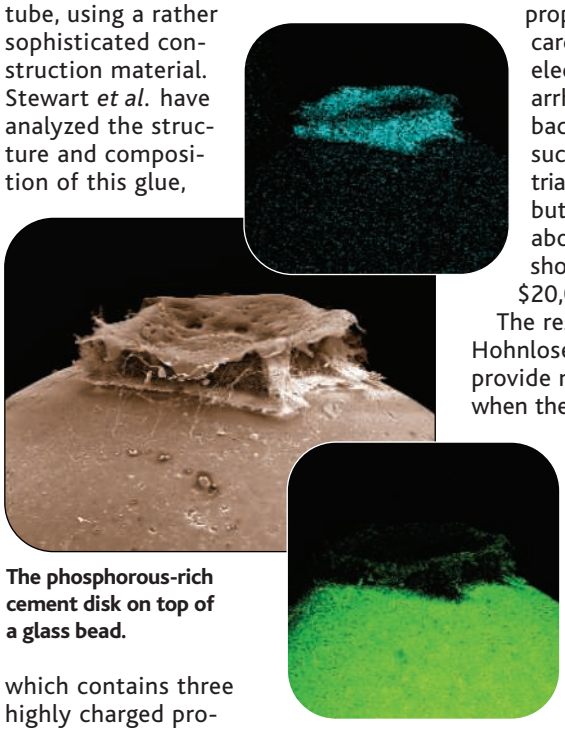
that this is not quite so, because (i) the number of fatalities per year is increasing; (ii) extreme events are not considered in the analyses; and (iii) the greatest seismic hazards and largest number of historic fatalities are concentrated in five countries: China, Iran, Italy, Japan, and Turkey, such that averaging over the global population tends to minimize the real problems. Today, there are about 100 cities of more than 3 million people, and half of these lie in earthquake zones. Soon, more people will live in cities than in rural areas, and by 2030 the population of Tokyo is predicted to reach 70 million. Combining the concentration of people in larger cities with the faster pace of construction caused by rapid growth means that it will be imperative to improve building codes and to monitor compliance more stringently in order to reduce earthquake fatalities. — LR

Seismol. Res. Lett. **75**, 706 (2004).

BIOCHEMISTRY

Quick-Drying Foam

Sandcastle worms build shelters for themselves by gathering sand grains and gluing them together into a sturdy tube, using a rather sophisticated construction material. Stewart *et al.* have analyzed the structure and composition of this glue,



The phosphorous-rich cement disk on top of a glass bead.

CREDITS: STEWART *ET AL.*, *J. EXP. BIOL.* **207**, 4727 (2004)

which contains three highly charged proteins: two are basic, whereas the third, acidic component accounts for the 30 mol % of phosphoserine in the cement. Concentrating these proteins (along with Ca^{2+} and Mg^{2+} to neutralize charge) within low-pH secretory granules in the cement gland initiates a process of complex coacervation. Phase separation occurs, yielding an emulsion-like blend of dehydrated proteins and cations along with water-rich droplets. When this mixture is daubed onto a sand grain, several changes occur, due in part to the higher pH and different ionic composition of seawater. The cation-phosphate interactions become ionic or salt-like in character, and the solvation of charges acts to soak up water from the cement/sand interface, improving contact as the cement sets. The hardened cement displays a cellular foam morphology, reflecting the separated phases, which also confers benefits in terms of an economy of material and a gradient of elasticity ideally suited to life in the intertidal zone. — GJC

J. Exp. Biol. **207**, 4727 (2004).

MEDICINE

Fighting Arrhythmias

People who have suffered a heart attack have a high risk of developing life-threatening arrhythmias. Because drugs do not effectively reduce this risk, there has been increasing interest in the prophylactic use of implantable cardioverter defibrillators (ICDs): electronic devices that detect arrhythmias and shock the heart back to its normal rhythm. The success of ICDs in early clinical trials has been a cause for optimism but has also prompted debate about how widely these devices should be used, given their cost ($20,000 each).

The results of a clinical trial by Hohnloser *et al.* suggest that ICDs provide much less benefit to patients when they are implanted within 6 weeks of a heart attack, as opposed to months or years later. Based on the results of a meta-analysis, Desai *et al.* conclude that ICDs can significantly increase the survival of a different group of patients—those who have a high risk of cardiac arrhythmias because of a heart condition called nonischemic cardiomyopathy. Together, these results emphasize the need for more extensive studies to define the patient populations most likely to benefit from these devices. — PAK

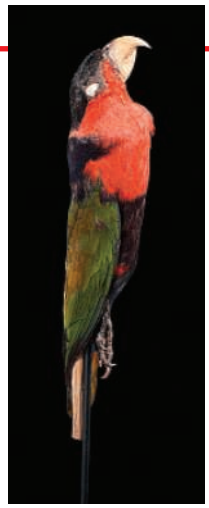
N. Engl. J. Med. **351**, 2481 (2004); *JAMA* **292**, 2874 (2004).

NetWatch

edited by Mitch Leslie

IMAGES

A Bird in Hand



For a nifty take on how museum collections can benefit from cyberspace, check out this digital specimen case from the Zoological Museum Amsterdam in the Netherlands. The site supplies three-dimensional (3-D) images of 151 avian type specimens from around the world—the original examples taxonomists used to describe the species. You can rotate or tilt animals ranging from crows and owls to this black-capped lory (*Lorius lory viridicrissalis*, right) from Indonesia. The pages also describe where and when the birds were collected, provide their measurements, and compare them to other specimens. The museum plans to post similar 3-D images of its cache of shells and skulls.

www.science.uva.nl/ZMA/3dpics

EDUCATION

Fire Up the Virtual Bunsen Burner

Demonstrating chemical reactions in class is a great way to spark students' interest—assuming the procedures work, everyone can see the lab bench, and nobody gets hurt. An alternative that eliminates these potential problems is this library of some 200 experiments for undergraduate labs from the Swiss Federal Institute of Technology in Zurich. You can search the experiment list by topic, keyword, or element to find everything from instructions for identifying metals by burning them to the synthesis of nylon. Movies of the reactions highlight important chemical transformations. Other features include a synopsis of the reaction, still photos of stages in the procedure, safety precautions, and references. Although some descriptions are in German, most experiments include English translations.

www.cci.ethz.ch/index.html

TOOLS

Sifting Through SNPs

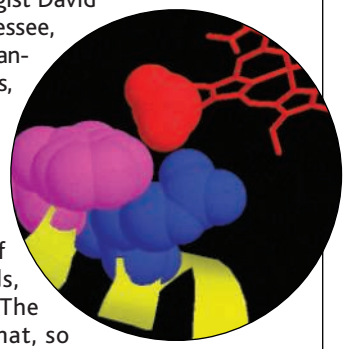
Researchers trawling for SNPs, or single-letter changes in the DNA code that might signal vulnerability to ailments such as cancer and heart disease, have a new tool to speed their search. The Ensembl human genome browser from the European Bioinformatics Institute now lets you chart how often particular SNPs travel together. Known as haplotypes, these patterns can help researchers choose the most informative SNPs to study. Access the feature, which lets you view data from several human populations, by searching for particular SNPs.

www.ensembl.org

DATABASE

Cytochrome Central

People who inherit a particular version of the gene *CYP2D6* don't get help from standard doses of the pain reliever codeine and can suffer side effects from many other medications. The problem is a sluggish drug-detoxifying enzyme from the cytochrome P450 family. This database from molecular biologist David Nelson of the University of Tennessee, Memphis, can help researchers get a handle on this sprawling group of enzymes, which take part in everything from breaking down Prozac and caffeine to synthesizing cholesterol.



The site lists more than 4000 versions of cytochrome P450 enzymes gleaned from published genomes of humans, honeybees, slime molds, bacteria, and other creatures. The sequences come in standard format, so you can plug them directly into genome analysis software or compare your sequences to those already on the site. For more information about cytochrome P450s, check out transcripts of Nelson's lectures or take a guided tour of some P450 molecules (above, *CYP2C5*).

drnelson.utmem.edu/CytochromeP450.html

EXHIBIT

The Making of the Atomic Bond

When Linus Pauling (1901-1994) was an undergraduate in chemistry, he began doubting the then-current notion that bonds form when tiny hooks on one atom slip into eyes on another. Pauling would go on to revolutionize our understanding of how atoms link up by sharing electrons, winning the Nobel Prize in 1954. A new site from Oregon State University in Corvallis, Pauling's alma mater, recounts this intellectual odyssey.

Pauling startled chemists in 1928 by announcing that he could use the new field of quantum mechanics to explain the long-standing question of why a carbon atom with four bonds forms a pyramid shape. You can browse the manuscript he published 3 years later that lays out his solution, listing six rules that describe electron sharing by atoms. The site includes other key publications—by the early 1930s, Pauling was writing a significant paper about every 5 weeks—along with stacks of photos, letters, and other memorabilia.

osulibrary.orst.edu/specialcollections/coll/pauling/bond/index.html

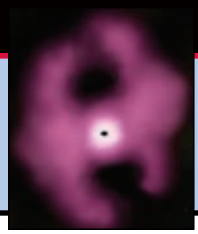
Send site suggestions to netwatch@aaas.org. Archive: www.sciencemag.org/netwatch

CREDITS (TOP TO BOTTOM): HUUB VELDHUIJZEN VAN ZANTEN/ETI; DAVID NELSON/UNIVERSITY OF TENNESSEE, MEMPHIS; OREGON STATE UNIVERSITY LIBRARIES/AVA HELEN AND LINUS PAULING PAPERS

Published by AAAS

This Week

PAGE 25
Getting religion fast



26
Gorging on galaxies

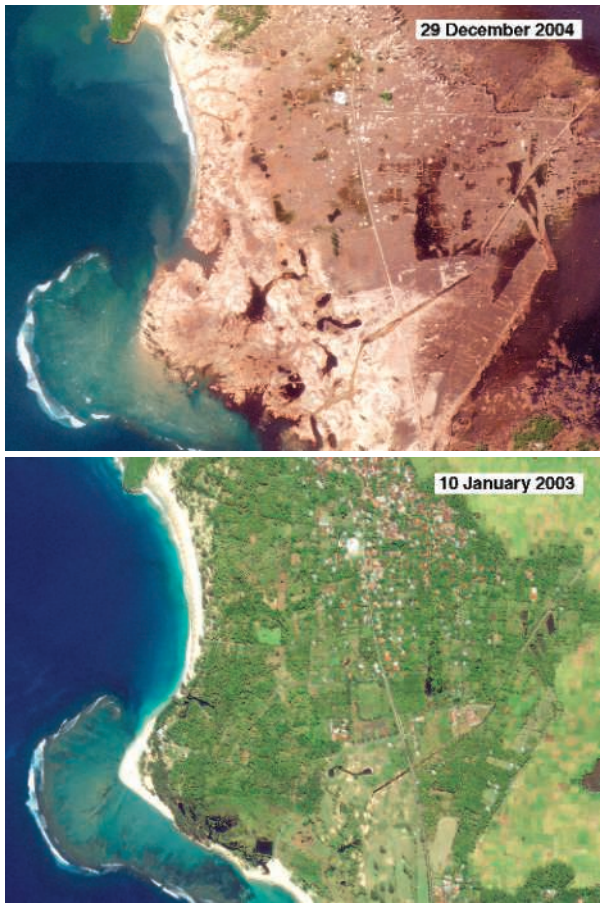
INDIAN OCEAN TSUNAMI

In Wake of Disaster, Scientists Seek Out Clues to Prevention

Having claimed more than 150,000 lives and destroyed billions of dollars' worth of property, nature last week reminded the world of the terrible cost of ignorance. Now the nations devastated by the massive earthquake and tsunami that ravaged the Bay of Bengal the morning after Christmas Day are hoping to marshal the political and scientific will to reduce the toll from the next natural disaster.

A week after the tragedy, the question of how many lives might have been saved had authorities in those countries recognized the danger in time to evacuate their coasts remains unanswered. But it's a hypothetical question, because the information needed to take such steps doesn't exist. That's why researchers are gearing up for an international data-collection effort in the affected countries, aimed at improving models of how tsunamis form and setting up a warning system in the Indian Ocean. "This was a momentous event both in human and scientific terms," says Costas Synolakis, a civil engineer and tsunami researcher at the University of Southern California in Los Angeles. "It was a failure of the entire hazards-mitigation community."

As relief efforts continue, scientists are traveling to the ravaged coasts to survey how far inland the water ran up at different points along the shorelines, how tall the waves were, and how fast they hit. In addition to providing a detailed picture of the event, says Philip Liu, a tsunami expert at Cornell University who is flying to Sri Lanka this week, information from these field surveys will enable researchers to test computer models that simulate the propagation of tsunami waves and the pattern of flooding when they break upon the shore. The geographical span of the disaster presents an opportunity to "run simulations on a scale that has not been possible with data from smaller tsunamis in the Pacific," says Synolakis, who is joining Liu in Sri Lanka.



Surprise attack. While tsunami waves ravaged towns such as Lhoknga, Indonesia (as shown in before-and-after satellite photos), scientists across the Bay of Bengal saw no danger coming.

Among other surveys being conducted in the region is one led by Hideo Matsutomi, a coastal engineer at Japan's Akita University, who is studying the disaster's effects on Thailand's shoreline.

Testing and refining tsunami models would increase their power to predict future events—not just in the Indian Ocean but elsewhere, too, says Vasily Titov, an applied mathematician and tsunami modeler at the Pacific Marine Environmental Laboratory in Seattle, Washington. Synolakis says the goal is to be able to predict, for any given coast with a given topography, which areas are most vulnerable and thus in greatest need of evacuation.

Such predictions would be easier to make if ocean basins resembled swimming pools and continents were rectangular-shaped slabs with perfect edges. But the uneven contours of sea floors and the jagged geometry of coastlines make tsunami modeling a complex engineering problem in the real world, Titov says. Exactly how a tsunami will travel through the ocean depends on factors including the intensity of the earthquake and the shape of the basin; how the waves will hit depends, among other factors, on the lay of the land at the shore.

What makes tsunami warnings even more complicated, Synolakis says, is that undersea quakes of magnitudes as great as 7.5 can often fail to generate tsunami waves taller than 5 centimeters. "What do you do without knowing precisely where and when the waves will strike and if they will be tall enough to be a threat?" he says. "Do you just scare tourists off the beach, and if nothing comes in, say, 'Oh, sorry'?"

It wasn't concerns about issuing a false alarm, however, that prevented scientists in India, Sri Lanka, and the Maldives from alerting authorities to the tsunami threat. Instead, researchers say, the reason was near-total ignorance. At the National Geophysical Research Institute (NGRI) in the south Indian city of Hyderabad, for example, seismologists knew of the earthquake within minutes after it struck but didn't consider the possibility of a tsunami until it was too late. In fact, at about 8 a.m., an hour after the tsunami had already begun its assault on Indian territory by pummeling the islands of Andaman and Nicobar some 200 km northwest of the epicenter, institute officials were reassuring the media that the Sumatran event posed no threat to the Indian subcontinent.

About the same time, in neighboring Sri Lanka, scientists at the country's only seismic monitoring station, in Kandy, reached a similar conclusion. "We knew that a quake had occurred—but on the other side of the ocean," says Sarath Weerawarnakula, ▶

Published by AAAS

director of Sri Lanka's Geological Survey and Mines Bureau, who hurried to his office that morning after feeling the tremors himself. "It wasn't supposed to affect us."

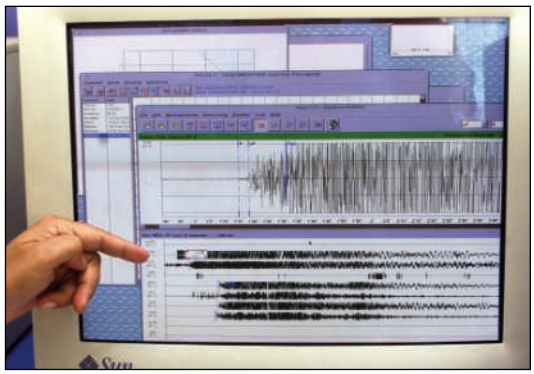
Walls of water crashing onto the Indian and Sri Lankan coasts soon proved how wrong the scientists were. The waves flung cars and trucks around like toys in a bathtub and rammed fishing boats into people's living rooms. "We'd never experienced anything like this before," says NGRI seismologist Rajender Chadha. "It took us completely by surprise, and it was a terrible feeling."

The international scientific community fared somewhat better at reacting to the quake, but not enough to make a difference. An hour after the quake, the Pacific Tsunami Warning Center (PTWC) in Ewa Beach, Hawaii—which serves a network of 26 countries in the Pacific basin, including Indonesia and Thailand—issued a bulletin identifying the possibility of a tsunami near the epicenter. But in the absence of real-time data from the Indian Ocean, which lacks the deep-sea pressure sensors and tide gauges that can spot tsunami waves at sea, PTWC officials "could not confirm that a tsunami had been generated," says Laura Kong, director of the International Tsunami Information Center in Honolulu, which works with PTWC to help countries in the Pacific deal with tsunami threats.

However, some researchers say that the seismic information alone—including magnitude, location, and estimated length of the fault line—should have set alarm bells ringing. Although not all undersea quakes produce life-threatening tsunamis, the Sumatran quake—later pegged at magnitude 9.0—was "so high on the scale, you had to know that a large tsunami would follow," says Emile Okal, a seismologist at Northwestern University in Evanston, Illinois. What may have made it difficult for officials to reach that conclusion, says Okal, was the rarity of tsunamis in the Indian Ocean: Fewer than half a dozen big ones have been recorded in the past 250 years.

But even if there had been reasonable certainty that a tsunami was building up stealthily under the waters, scientists say they are not sure what they could have done. As the morning wore on, for example, geophysicists in India realized that "a tsunami would be generated, but how it would travel and when it would strike—we simply had no clue," says Chadha.

That's exactly the kind of information that countries in the region hope to have the next time a tsunami comes calling. The Indian government last week announced plans to spend $30 million to set up a warning system within the next 2 years; Indonesia and Thailand have since announced similar plans of their own. Like those in the Pacific, the proposed warning systems will include up to a dozen deep-sea buoys to detect pressure changes that occur as an earthquake's energy travels through the ocean and tide gauges to measure rise and fall in sea level.



Off the scale. The Sumatra quake turned out to be far more powerful than early readings suggested.

Kapil Sibal, minister of state for science and technology and ocean development, says India plans to collaborate with Indonesia, Thailand, and Myanmar to eventually build a tsunami warning network in the region. "We've been jolted hard, and we'll take remedial action," Sibal says.

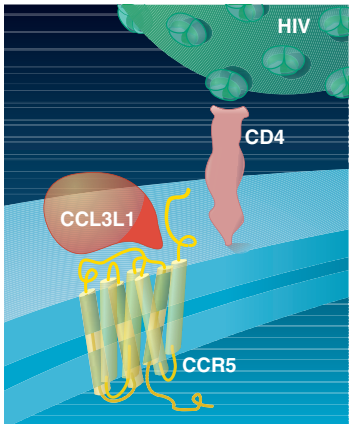
–YUDHIJIT BHATTACHARJEE

With reporting by Pallava Bagla in New Delhi.

VIROLOGY

Chemokine Gene Number Tied to HIV Susceptibility, But With a Twist

Like a long-married couple, a virus and its host shape each other in subtle yet profound ways. AIDS researchers investigating this dynamic have detected several changes in both HIV and humans that likely evolved during the high-stakes wrestling match between the virus, the cells it infects, and the immune system. Now a massive review of DNA from more than 5000 HIV-infected and uninfected people has found that the human genome appears to have responded to the virus by stockpiling extra copies of immune genes that influence a person's HIV susceptibility as well as the course of disease in infected people. These findings may lead to an important practical advance: better designed AIDS vaccine studies.



No vacancy. When *CCL3L1* (red) occupies the CCR5 receptor on CD4 cells, it blocks HIV's entry.

Described in the 6 January *Science* Express (www.sciencemag.org/cgi/content/abstract/1101160), the DNA analysis focuses on a gene with the ungainly name of *CCL3L1*. Steven Wolinsky, a virologist at Northwestern University Medical School in Chicago, Illinois, whose lab also has studied the relationship between immune genes and HIV, calls the work "an intellectual and technical tour de force."

Sunil Ahuja, an infectious-disease specialist at the Veterans Administration Research Center for AIDS and HIV-1 Infection in San Antonio, Texas, led an international team that examined the importance of segmental duplications in the human genome. People typically have two copies of each gene (one from each parent), but stretches of DNA sometimes appear repeatedly, causing the overrepresentation of certain genes. Many of the segmental duplications discovered to date include genes related to immunity, inspiring the notion that some duplications protect against invaders such as viruses. Ahuja and co-workers wondered whether HIV might be the target of such an evolutionary response.

The researchers first hunted for segmental duplications that include *CCL3L1* in 1000 people from 57 populations. Immune ▶

CREDITS (TOP TO BOTTOM): TATAN SYUFLANA/AP PHOTO; K. SUTLIFF/*SCIENCE*

Published by AAAS

cells signal one another using chemicals called chemokines, and *CCL3L1* codes for one that docks onto the same white blood cell receptor, CCR5, that HIV grabs to infect the cells. In theory, as levels of this chemokine rise, it fills more CCR5 receptors, blocking HIV's ability to infect.

Ahuja and his colleagues found that the copy number of *CCL3L1* varies from person to person and influences an individual's level of the chemokine. But by itself, this number didn't determine HIV susceptibility. Rather, it depended on how many copies a person had compared to others of the same ancestry. For example, their review revealed that Africans had a median of four copies of *CCL3L1*, whereas Europeans had an average of two. At first blush, this evidence seems to suggest that HIV might have a more difficult time causing harm in Africans. But a closer analysis revealed nothing of the sort.

The U.S. military for 20 years has closely followed a racially diverse cohort of HIV-infected people. Ahuja joined a team led by Matthew Dolan of the Tri-Service AIDS Clinical Consortium to use DNA from these 1000 people to help unravel the relation between *CCL3L1* and HIV. After matching the cohort by race and ethnicity to more than 2000 uninfected controls, the researchers compared how many copies of *CCL3L1* each person had. From these data, they concluded that segmental duplications of the gene thwarted infection in the controls and slowed disease in the infected—but only if people had a higher number than average for their racial or ethnic background. And people who had fewer copies of the gene relative to members of their ethnic group—including babies of infected mothers—had increased susceptibility to HIV.

Factoring in *CCL3L1* status could help separate wheat from chaff in AIDS vaccine studies. To date, vaccine testers have paid little attention to differences in genetic susceptibility to HIV. But if a person has, say, a high level of genetic protection, a vaccine might appear to work when it did not. Conversely, highly susceptible people could make a good vaccine look bad. Ahuja and co-workers propose that by analyzing *CCL3L1* and similar genetic factors together, researchers could illuminate the now invisible line that separates the effects of vaccines from the power of the host's genes. **–JON COHEN**

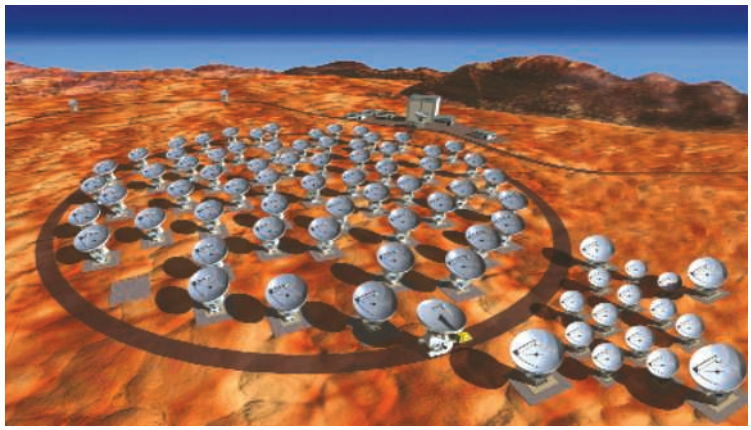
JAPAN

New Budget Accelerates Shift to Competitive Grants

TOKYO—Academic research in Japan appears to have more than held its own in a tight funding year. A 2005 budget adopted last week by the cabinet of Prime Minister Junichiro Koizumi features a 2.6% boost for the direct funding of research, far outpacing a 0.1% rise in overall government spending. It also bucks a 0.8% dip in the country's total science budget, the first such decline in decades. "Given how tight the government budget is, this is not so bad," says Akio Yuki, vice minister of the Ministry of Education, which accounts for the bulk of Japan's scientific efforts.

The decline in science-related spending overall, to $34.1 billion in the fiscal year that starts 1 April, is driven by a 22% decrease in defense research and development. The chief cuts are in new weapons systems and aircraft procurement. Most of this money goes to defense contractors, however, and "has little connection to academic research," says Reiko Kuroda, a biochemist at the University of Tokyo and a member of the Council for Science and Technology Policy, the nation's highest science advisory body. The government also fell short of its 2000 promise to double science spending over 5 years, to an aggregate 24 trillion yen ($229 billion). Officials blame a sluggish economy, although they expect government spending to reach 75% of that goal by the end of the fiscal year.

The $12.6 billion slated for day-to-day research needs such as supplies and equipment includes a 30% rise in funding for competitive grants, to $4.4 billion. That's part of a concerted effort to wean university scientists off a system of small but universal block grants and onto one that rewards the best ideas. The increased support, up 57% since 2000, comes from a combination of new funding and a diversion of resources from older, directed programs in fields such as nuclear power engineering. "There was a lot of resistance," Kuroda says about the shift to a more open process (*Science*, 27 June 2003, p. 2027). But she says that Koizumi, the nominal head of the science council, applied the political pressure needed to bring the bureaucrats in line.



Tuning in. Japan will more than double funding this year for the Atacama Large Millimeter/Submillimeter Array in Chile, a joint project under way with the United States and Europe.

CREDIT: NATIONAL ASTRONOMICAL OBSERVATORY OF JAPAN

Universities will also feel the bite of increased competition. The new budget allows them for the first time to claim 30% of selected large grants for administrative costs and overhead. In return, however, the government is cutting back on a fund that supports operating expenses on campus. The bottom line is that universities will become more dependent for their operating expenses on grants to individual researchers, a change that Kuroda and others worry could have a negative impact on institutions that put a greater emphasis on teaching than on research.

There's good news for universities funded by the Ministry of Education, where science funding is rising almost across the board. In addition to competitive grants, areas receiving significant boosts include big-ticket facilities, such as the Atacama Large Millimeter/Submillimeter Array being built in Chile, and projects expected to have a short-term economic payoff. Favored fields include the life and environmental sciences, nanotechnology, and information technology.

The science council has not yet settled on spending targets for a third 5-year plan that would run through the 2010 fiscal year. But the business community is already lobbying for continued increases in science. In November, the Keidanren, Japan's most influential business group, called on the government to hold firm to its goal of raising science spending to 1% of the country's gross domestic product. That percentage is expected to stand at 0.8% by the end of the 2005 fiscal year. "The industrial sector has had to cut back on basic R&D," says Keiichi Nagamatsu, Keidanren's managing director. "We're looking to the universities to fill that role."

The cabinet adopted the 2005 budget on 24 December. It now goes to the Diet, Japan's legislative branch, where approval is typically routine. **–DENNIS NORMILE**

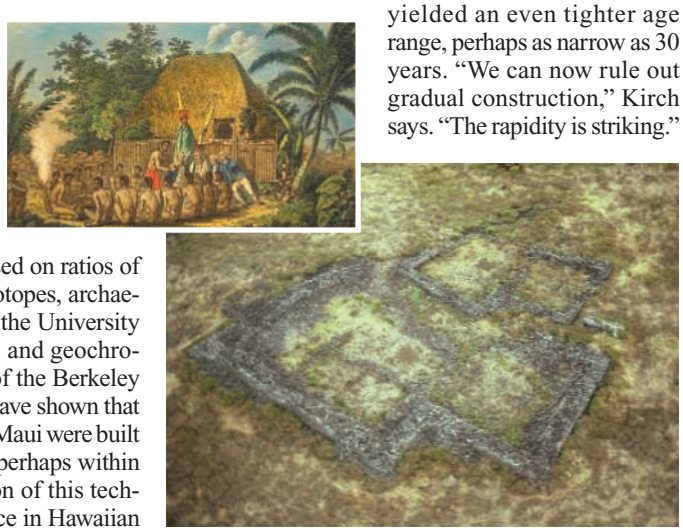
Published by AAAS

ANTHROPOLOGY

Coral Ages Show Hawaiian Temples Sprang From Political Revolution

Hawaiian legends say a ruler named Pi'ilani brought peace to Maui by routing rival chiefs, marrying a powerful queen, and setting himself up as absolute ruler. Historians agree that this progression from feuding chiefs to kingdom, repeated on several other of the Hawaiian Islands, ultimately created a highly stratified society with elaborate religious rituals that justified the divine right of kings. But they have never been sure how long it took for a religious state to emerge.

Now a preliminary study of temples on Maui, described on page 102 of this issue of *Science*, suggests it may have happened within a single generation, around 1600 C.E., just as the stories suggest. By dating coral offerings using a geological technique based on ratios of uranium and thorium isotopes, archaeologist Patrick Kirch of the University of California, Berkeley, and geochronologist Warren Sharp of the Berkeley Geochronology Center have shown that several large temples on Maui were built at about the same time, perhaps within 30 years. The application of this technique is "a major advance in Hawaiian archaeology," says J. Stephen Athens of the International Archaeological Research Institute Inc. in Honolulu.



Power base. Ruins on Maui suggest that the island's first king exerted control by quickly building temples, such as those seen elsewhere by Captain Cook (*inset*).

The most sophisticated and stratified societies in the Pacific evolved on the Hawaiian Islands. Oral histories written down in the 19th century provide a rich source of information about the rise of royalty. Other clues come from the many temples these rulers built to demonstrate their divine power and to receive tribute. Yet the technique normally used to measure ancient artifacts, radiocarbon dating, can't get a clear fix on such recent history.

Kirch and Sharp solved that problem by applying another kind of radiometric dating typically used to date high-and-dry coral reefs and reconstruct the history of sea level. When Hawaiians built temples to agricultural gods, they placed coral into the basalt walls and foundations, presumably as offerings. Because the coral preserves fine details, Kirch and Sharp argue that it was freshly cut from living reefs. By dating the coral, they could find out when the temples were constructed.

As coral-producing organisms grow, they incorporate uranium atoms in seawater into their skeletons. The uranium atoms decay into thorium-230 at a precisely known rate. So by measuring the ratio of uranium-238 to thorium-230, the researchers could tell precisely how long ago the coral had been cut from the reef.

To their surprise, samples from eight temples on southeast Maui, including one as large as 1400 square meters (see photo), all yielded dates between 1580 and 1640 C.E. The samples that most accurately reflected the time of collection from the sea—those from the tips of branches, the youngest part of the coral—yielded an even tighter age range, perhaps as narrow as 30 years. "We can now rule out gradual construction," Kirch says. "The rapidity is striking."

That fast pace, Kirch and Sharp argue, implies a major change in politics. "It looks like one person taking control of the system and ratcheting up [his power]," Kirch says, because only a powerful ruler could have marshaled the labor to build such temples so quickly. Michael Kolb of Northern Illinois University in DeKalb suggests that the similarity of the offerings could also indicate a centralized authority. "The standardization of worship hints at state religion," he says. "It shows you just how centralized the power was."

The ruler could very well have been Pi'ilani, Kirch and Sharp say. A count of generations in the oral histories suggests that he reigned from roughly 1570 to 1600 C.E. Once Maui had been unified, however, Pi'ilani's peace didn't last. Descendents began to fight the kings of other islands in ever-bloodier battles. Interisland warfare lasted until Kamehameha the Great of Hawaii consolidated power in 1805 through the use of weapons obtained from the Europeans.

–ERIK STOKSTAD

CREDITS: P. V. KIRCH; (INSET) BY PERMISSION OF THE NATIONAL LIBRARY OF AUSTRALIA

ScienceScope

Although history, 2004 is casting a shadow. This year researchers will continue to deal with the fallout from conflicts over consulting, clinical trials oversight, and other issues.

Consulting Turmoil

A controversy over industry consulting by staff scientists will likely loom over the National Institutes of Health (NIH) well into 2005, possibly hindering efforts to retain and attract top talent. The fate of the agency's $28.4 billion budget could rest on Director Elias Zerhouni's ability to satisfy critics without alienating staff.

The uproar began in late 2003 when the *Los Angeles Times* reported that several scientists at NIH had received hundreds of thousands of dollars in payments from drug companies, sparking a congressional investigation (*Science*, 19 December 2003, p. 2046). Last month, the newspaper alleged that other prominent researchers improperly consulted for drug or product manufacturers on topics that involved their official work. The paper's editors called for Zerhouni's resignation, but he fired back with a letter denying "complacency" and defending NIH's "new stringent rules," which include a 1-year ban on all industry consulting and limits on lecture honoraria.

Meanwhile, those proposed rules have angered many agency scientists. They have also hindered recruitment of intramural directors for the neurological disorders and mental health institutes, sources suggest. And scrutiny has contributed to at least one departure: Alzheimer's researcher Trey Sunderland, who reportedly didn't disclose to NIH ethics officials some of his consulting activities, is leaving for the Albert Einstein College of Medicine in New York City.

–JOCELYN KAISER

Trials by Fire

The U.S. Food and Drug Administration (FDA) faces a push this year from federal legislators bent on overhauling how the agency monitors drug safety. One vehicle may be a bill creating a mandatory clinical trials registry, an idea that picked up steam last year after the pharmaceutical industry and FDA ran into sharp criticism for their handling of antidepressants linked to suicidal behavior in children and teenagers. Adding fuel to the fire is the ongoing debate over harmful side effects from anti-inflammatory pain medications, such as COX-2 inhibitors. How these proposals will fare is unclear. The Republican-led Congress, the White House, and the pharmaceutical industry have traditionally shied from hands-on drug monitoring. But a frustrated and confused public may demand greater regulation.

–JENNIFER COUZIN

Published by AAAS

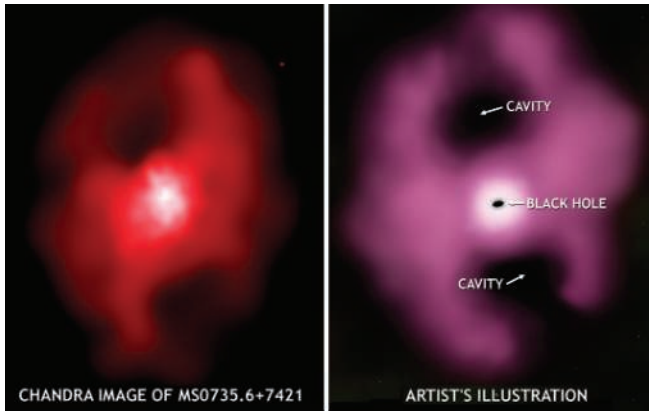
ASTROPHYSICS

Gorging Black Hole Carves Out Gigantic Cavities of Gas

The most energetic eruption yet found in space has yielded the first direct measure of a black hole's prodigious appetite. The outburst, still going strong after 100 million years, has gouged two enormous cavities within the hot gas in a distant cluster of galaxies. The stark features show that even mature black holes can disrupt star birth and influence matter far beyond their host galaxies.

Each of the "supercavities," reported in the 6 January issue of *Nature*, could swallow 600 galaxies the size of our Milky Way. To shove aside such vast volumes of gas, the eruption has churned out as much energy as nearly a billion gamma-ray bursts—the most powerful impulsive explosions known. "Seeing this huge amount of energy was quite surprising, one might even say shocking," says astrophysicist Richard Mushotzky of NASA's Goddard Space Flight Center in Greenbelt, Maryland, who is not part of the research team.

The cavities appear in a galactic group called MS0735.6+7421, about 2.6 billion light-years from Earth. The fully developed cluster looks unremarkable in visible light, says the study's lead author, astronomer Brian McNamara of Ohio University in Athens. At its center resides a supermassive galaxy, bloated by billions of years of consuming smaller galaxies in the cluster. Radio images had revealed a classic double-sided jet of energy streaming away from this central galaxy, suggesting that it hosts a black hole still gorging on infalling gas.



Gaping holes. X-rays from hot gas in a cluster of galaxies (left) outline two "supercavities" cleared out by an eruption from a central black hole (artist's view, right).

An 11-hour observation by NASA's Chandra X-ray Observatory exposed voids in the hot gas that pervades the cluster, cleared out along the paths of the radio jets. By tracing the sizes of those voids, the astronomers measured how hard the black hole had to work to displace the gas—in the same way that lungs need to exert more force to inflate a larger balloon. "[The supercavities] allow us to measure the energy deposited by the central black hole into its surroundings in the most direct possible fashion," McNamara says.

The calculation shows that the black hole must have devoured about three times the mass of our sun each year for the last 100 million years, says co-author Paul Nulsen of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts. That average rate is similar to the feeding frenzy that probably powered quasars at the cores of galaxies in the early universe, but it's unheard of in modern galaxies. Thus, it appears that black holes within some clusters may have grown at a fantastic rate even in relatively recent times, Nulsen says.

The eruption also gives tangible evidence of a poorly understood process that helps shape how the cosmos looks, Mushotzky notes. Astrophysicists have long suspected that "feedback" of blazing energy from the centers of galaxies can heat gas for millions to billions of years, preventing new stars from forming as quickly as models predict. The details are still elusive, but the new work offers some insights. "Here, for the first time, you're actually seeing the energy injected and the gas being heated," Mushotzky says. "We're all really excited."

—ROBERT IRION

PHYSICS

Cesium Collisions Help Create Colder Antihydrogen

A clever new way to make antihydrogen may bring scientists one step closer to understanding how matter differs from antimatter.

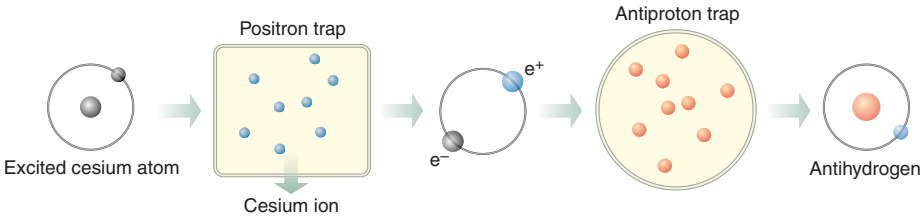
In the 31 December issue of *Physical Review Letters*, a group of physicists describes a laser-assisted technique to make antihydrogen, which mirrors everyday hydrogen by consisting of an antiproton bound to an antielectron. "It's really very different in principle" from previous methods of making antihydrogen, says Gerald Gabrielse, a physicist at Harvard University, who worked with a handful of antimatter-makers known as the ATRAP collaboration to develop the new approach.

For years, ATRAP and a rival group, ATHENA, have been cooling antiprotons (which come from a beam at CERN, the European particle physics lab near Geneva, Switzerland) and antielectrons (which come from a radioactive source) and mixing them in a magnetic bottle in hopes of producing antihydrogen. Both teams have created thousands of antihydrogen atoms this way (*Science*, 15 November 2002, p. 1327). However, those antihydrogens were relatively warm—several degrees above absolute zero—and, therefore, moving too fast to capture and study in detail.

ATRAP's new method collects antiprotons and antielectrons in separate magnetic traps. Then the researchers shoot atoms toward the antielectrons, exciting the atoms with lasers to force their electrons into larger-than-typical orbits around the nucleus. "We make [the cesium atoms] very big—and a big thing has a higher probability" of striking an antielectron in the trap, says Gabrielse.

After impact, the cesium's electron binds to the antielectron, forming an unstable and excited conglomerate known as positronium. The positroniums zoom away in all directions, and some wind up in the nearby trap containing antiprotons. Following another collision, the antielectron once again jumps ship and hops to the antiproton, forming an excited antihydrogen.

This Rube Goldberg–ish method has so far produced fewer than two dozen anti- ▶



Ave, cesium. A beam of excited cesium atoms hits a trap full of antielectrons, and the products fall into a pile of antiprotons. The result: cold antihydrogens.

CREDITS (TOP TO BOTTOM): NASA/CXC/OHIO UNIVERSITY/B. MCNAMARA *ET AL.*; ADAPTED FROM C. H. STORRY *ET AL.*, *PRL* **93**, 263401 (2004)

Published by AAAS

hydrogens. But in principle, it should allow physicists to create very cold and slow-moving antihydrogens. "Since the positronium is so lightweight compared to the antiprotons, when they collide, it's very hard for them to heat up the antiprotons," says Gabrielse. Because physicists can potentially cool antiprotons to within a few hundred thousandths of a degree of absolute zero, this method might, without too much tweaking, yield antihydrogens slow enough to study.

"Anything that goes in this direction is welcome," says Rolf Landua, a CERN physicist and member of the ATHENA collaboration. But the low yield is a problem, he cautions, and studying the produced antihydrogen properly will likely require deexciting the atoms, perhaps with another laser. "Maybe, in the end, that will be the way forward, but it looks complicated," Landua says.

Unfortunately, scientists will have to wait to find out. The antiproton source at CERN has been shut down until 2006 to speed construction of the Large Hadron Collider.

–CHARLES SEIFE

AVIAN FLU

Mild Illnesses Confound Researchers

TOKYO—Ten months after an outbreak of highly pathogenic avian influenza, researchers in Japan have confirmed that four employees of an infected farm and one governmental health official are carrying antibodies to the H5N1 virus. These are the first documented cases of mild or asymptomatic infections in humans to emerge from last year's outbreak. In Vietnam and Thailand, the disease resulted in death in more than 70% of confirmed human cases.

Viruses "typically" cause a wide range of symptoms in humans, says Yi Guan of the University of Hong Kong, who has studied H5N1 since it emerged there in 1997. Similar results were found in surveys of wild-animal dealers in China after the 2002 severe acute respiratory syndrome outbreak and among cullers and poultry workers in Hong Kong after the 1997 H5N1 outbreak. The new cases should help scientists understand the behavior of avian flu in humans. "It is important to learn what percentage of people exposed to the virus become infected, and among those, how many develop severe and how many develop mild illnesses," he adds.

When the Japanese H5N1 outbreak was confirmed at a chicken farm in Kyoto Prefecture last February, Japan's National Institute of Infectious Diseases urged local officials to survey farm workers, health inspectors, and those who destroyed the chickens. Institute virologist Masato Tashiro, director of the World Health Organization collaborative center for influenza surveillance and research in Japan, says the difficulties in detecting low levels of antibodies slowed the work, and then prefectural officials dithered over releasing the results.

Out of 7000 people potentially exposed, only 58 agreed to participate in the survey. Those 58 included 17 of 19 people who worked on the infected farm before taking the antiviral medication Tamiflu or wearing protective clothing. The five people who proved to be seropositive were among this group; none of those who took Tamiflu before going to the farm or wore protective gear while there proved positive. "We think this does say something about the value of antiviral medication and proper protection," notes Tashiro.

Spot check. A worker draws chicken blood for disease testing.

CREDIT: ERIKO SUGITA/REUTERS

Albert Osterhaus, a virologist at the Netherlands' Erasmus University Medical Center in Rotterdam, suggests that the five Japanese could have developed antibodies in response to viral antigens in the farm environment and were never actually infected with the H5N1 virus.

Why the infections, if they did occur, proved so mild is less clear. Tashiro offers several possibilities. For one, the genetic sequences of the viral strain collected in Japan and Korea varies from that of the strain that appeared later in Thailand and Vietnam. Once the presence of H5N1 was confirmed, farm workers and health official who had visited the farm took Tamiflu, perhaps in time to reduce the severity of the infection. Finally, exposure to the virus could have been more limited than among the patients in Thailand and Vietnam, many of whom raised chickens at home.

"We don't have any controls, so it's difficult to determine just why these differences occurred," Tashiro says. Scientists hope that surveys of cullers in Thailand and Vietnam who did not take Tamiflu and were often not wearing proper protective gear may answer these questions.

–DENNIS NORMILE

With reporting by Martin Enserink.

ScienceScope

Space Program Shakeup

NASA soon faces some key scientific decisions and budget issues, starting with who will succeed Sean O'Keefe as administrator. The White House is likely to nominate a new chief in the next few weeks, and a Senate confirmation hearing could come as early as February, in time for the start of the 2006 budget battle.

The new space agency leader will have to wrestle with whether to service the aging Hubble Space Telescope with the shuttle—as astronomers prefer—or with a robotic mission. And he or she will have to persuade Congress to fund the moon-Mars human exploration effort proposed a year ago by President George W. Bush. To bring some budget discipline to that program, NASA Comptroller Steve Isakowitz, a longtime White House budget official, will take over as deputy in the exploration office. One of his tasks will be to decide whether a new nuclear propulsion system, dubbed Prometheus, should first be used to head for Jupiter's icy moons or Earth's moon.

–ANDREW LAWLER

Aux Barricades?

PARIS—In the wake of protests by researchers last year, the French government is expected to unveil a new bill next week to bolster the nation's sciences. Described as a reform package, it's intended to make scientific careers more attractive and improve the national funding and evaluation of research. But scientists say they fear it may go in the wrong direction.

Early signals about the plan "are not good," says Alain Trautmann, co-director of the cell biology department at the Cochin Institute and spokesperson for the protest movement last year that forced the government to back down on spending and job cuts (*Science*, 16 April 2004, p. 368). The biggest worry is about jobs. Leaders of the protest movement criticized the government just before Christmas for, among other things, announcing a "derisory" 150 new university lecturer-researcher posts in the 2005 budget. Hundreds more are needed, says Edouard Brézin, incoming president of the French Academy of Sciences, if the government is serious about reducing their teaching hours. If the bill falls short, researchers say, they will take to the barricades again. The legislation is expected to reach Parliament for a vote by summer.

–BARBARA CASASSUS

Published by AAAS

PLANETARY EXPLORATION

Europe Draws Up Its Own Strategy for Visiting the Moon and Mars

CAMBRIDGE, U.K.—President George W. Bush's announcement last January of a major push to explore the moon and Mars may have generated lots of headlines (*Science*, 16 January 2004, p. 293). But while the fate of that plan remains up in the air, Europe's own strategy for planetary exploration, begun 3 years ago, is gathering real support.

Late last month the European Space Agency (ESA) announced that member states had nearly tripled the budget for the Aurora program, which is planning a series of missions culminating in a crewed visit to Mars in 2033. Although many researchers are wary of the commitment to send astronauts, they generally support Aurora's aims. "As someone who is interested in planets, Aurora can do it for us," says John Zarnecki of the Open University in Milton Keynes, U.K.

First step. Plans call for a launch of ExoMars in 2011.

ESA's initial proposal for Aurora in 2001 attracted just $19 million of the $27 million requested from members—an inauspicious start. Piero Messina, Aurora spokesperson at ESA headquarters in Paris, says the shortfall occurred when Italy, a strong supporter of planetary exploration, suddenly had a change of government and "could not live up to its earlier commitments."

ESA researchers began work with what they had, but before long the context had changed. NASA's prolonged grounding of the shuttle fleet in February 2003 and a reduced U.S. commitment to the international space station created problems, whereas NASA's new moon-Mars program opened up new possibilities for collaboration. In July, ESA asked member states to provide new money for studies. Italy came through with $17 million on top of its original $3.4 million, currently making it the largest contributor to the $56 million that has been pledged. Another surprise was the additional $6.7 million from the U.K., a longtime opponent of crewed missions.

A mission strategy will be hammered out over the coming year. Messina says ESA researchers are working on three possible scenarios for lunar exploration that they will present next month. Aurora's first mission, the 2011 ExoMars orbiter and lander, is already well defined. ESA and NASA are both planning missions to bring samples back from Mars, and officials from both agencies are now working out how they might collaborate. "There is a will to converge," says Messina.

The pressure to cut costs will intensify by the year's end when ESA presents the full Aurora program to ministers from member nations. Messina estimates that ESA will need $1.3 billion for the first 5 years to begin building the spacecraft. Ian Halliday, head of the U.K. Particle Physics and Astronomy Research Council, says ESA's current cost projections are "wishful thinking." Even its supporters don't dare hazard a guess about its prospects. Says Zarnecki, "I haven't a clue."

—DANIEL CLERY

SYSTEMATICS

Philadelphia Institution Forced to Cut Curators

A chronic budget shortfall has forced the oldest natural history institution in the United States to lay off 5% of its staff. Outside scientists are especially concerned that the Academy of Natural Sciences in Philadelphia is losing three of its 10 curators, including the overseer of a prized, nearly 200-year-old ornithology collection. The move is part of a trend of cutbacks at natural history museums. "We're losing positions. It's of national concern," says Smithsonian Institution ornithologist Helen F. James.

Scientific treasure. The ornithology collection at the financially troubled Philadelphia academy includes specimens of the extinct Australian paradise parrot (*Psephotus pulcherrimus*).

The academy, founded in 1812, runs a museum and research programs and houses 17 million biological specimens. Its $12 million annual budget has faced deficits of $500,000 to $1 million for a decade, explains president and CEO D. James Baker, former head of the National Oceanic and Atmospheric Administration. As a result, Baker says leaders made the "painful decision" last month to lay off 13 of 250 employees across all divisions. The layoffs go into effect over the next 6 months. Thomas Lovejoy, head of the Heinz Center, an environmental think tank in Washington, D.C., and an academy board member, says that the cuts were inevitable. "They just had to address" the deficit, he notes.

The three curators losing their jobs are Leo Joseph, assistant curator and chair of ornithology; Richard McCourt, an associate botany curator; and Dominique Didier-Dagit, an associate curator of ichthyology. Some outside scientists who asked not to be identified suggest that these junior scientists weren't pulling in enough grant money. Baker doesn't deny the charge, saying that the academy tried to keep staff in "areas where we think there is research support from outside agencies." (Joseph and McCourt referred calls to an academy spokesperson.)

The academy's ornithology collection, which now has no curator, is a paramount concern. The holdings include many of the earliest specimens collected by North American ornithologists as well as the Australia collection of British ornithologist John Gould. Baker says the academy "has made an absolute commitment to preserve" this resource, which will still have a manager to make it available to scientists. But experts worry that the absence of a curator to add specimens and conduct his or her own research could undermine it. "A collection should be part of a living and breathing community," says A. Townsend Peterson, ornithology curator of the Natural History Museum at the University of Kansas, Lawrence.

Baker is mum on future staffing plans, saying only that "we can grow our number of curators" if the budget outlook improves. But he predicts that a focus on certain areas, such as watershed management and molecular systematics, will create "a stronger institution."

—JOCELYN KAISER

CREDITS (TOP TO BOTTOM): ESA; THE ACADEMY OF NATURAL SCIENCES

Published by AAAS

UNIVERSITY ASSESSMENT

Funding Woes Delay Survey of U.S. Graduate Programs

The National Research Council (NRC) is having trouble raising enough money for an assessment of U.S. doctoral programs. Everybody agrees that a survey of the quality of U.S. graduate education is important. But the consensus dissolves when it comes to paying for it.

The National Academies' NRC is trying to raise $5.2 million for what it hopes will be a bigger and better version of two previous assessments, which appeared in 1982 and 1995, of the relative quality of research doctoral programs. Two foundations—Alfred P. Sloan and Andrew W. Mellon—have agreed to kick in $1.2 million, roughly the cost of the 1995 survey. But NRC's attempt to collect the rest from the federal government has so far come up empty. "We've talked to many agencies, but we haven't generated any interest," laments one NRC official.

As a result, last month NRC officially postponed by 1 year the scheduled 1 July 2005 start of the assessment, a multistage exercise that includes a compilation of institution and program demographics, an analysis of each faculty member's publishing record, and a polling of graduate students. (An earlier schedule had the survey beginning last summer.) The decision, which study director Charlotte Kuh blames on "a delay in funding," means an expected publication date of 2008 rather than the original target of 2006.

That's a blow to what Princeton University astrophysicist Jeremiah Ostriker calls "the premier way to measure graduate education." Ostriker chaired an NRC panel whose recommendations on methodology and scope have been incorporated into the new survey (*Science*, 12 December 2003, p. 1883). The delay cedes ground to commercial rankings, notably by *U.S. News and World Report*. It also complicates life for U.S. institutions with aspiring programs that look to the NRC survey to validate their progress at a time when graduate schools are facing growing competition from other nations for the world's best students.

The holdup is a big disappointment to J. Bruce Rafert, dean of the graduate school at Clemson University in South Carolina, who persuaded his bosses to pony up additional resources to gather data from faculty, students, and staff to pass along to NRC. "I had coordinated data collection with the IT people and held a number of workshops for faculty and staff," says Rafert. "We were fairly far into this when I heard [about the delay]."

Some administrators aren't taking the news lying down. In a meeting last month of graduate deans, Lawrence Martin of Stony Brook University in New York proposed that universities pay an annual subscription fee to raise the necessary funds. "Of course the government has a stake," says Martin. "But if the feds don't want to pay, then we have to do it another way. For me, it's not an option not to do it." A modest annual fee, Martin noted, would also allow NRC to update the survey more frequently than the current rate of once every 13 years. The proposal makes a lot of sense to many deans. "It's the best suggestion that I heard at the meeting," says Rafert.

But other administrators are cool, if not downright hostile, to financing the survey that way. Universities would already be paying indirectly for the assessment with a sizable investment of staff time and resources, argues John Vaughn of the Association of American Universities in Washington, D.C., a coalition of 62 major research institutions in the United States and Canada. He also thinks the assessment will generate data that can help the federal government gauge the quality of the scientists whom it is supporting.

"I think [a subscription] would be a real mistake because graduate training is a society-wide issue," says Vaughn. "It's also a slippery slope; if universities pick up the tab for this, then the government may start looking to duck other obligations, too." Debra Stewart, president of the Washington, D.C.–based Council of Graduate Schools, also fears that the survey's credibility could be tainted if its primary audience also pays the freight.

Academy officials hope to meet this month with presidential science adviser John Marburger to make the case for the government's involvement. (Neither of the previous NRC surveys received federal funding, although the National Institutes of Health, the National Science Foundation, and the U.S. Department of Agriculture helped finance the methodology review that Ostriker chaired.) But they may need stronger arguments than those they've used to date.

"The NRC survey is well-designed and likely to be an improvement on all previous assessments," Marburger said in an e-mail to *Science*. "But it is more directly relevant and useful to the surveyed institutions than to the funding agencies." One government official who has heard NRC's pitch found it lacking. "We thought that we could use the technical portion of the assessment to help us evaluate our own training programs," says the official, who requested anonymity. "But that idea doesn't really hold up. We already get a lot of information from our grantees." At the same time, the official added, some issues of interest to an agency may be too specialized to show up in the NRC survey.

Although Vaughn sees NRC's suspension of the survey as a necessary evil, Martin worries that it could be the beginning of the end. "After telling people get ready, get ready for the NRC survey, now I'm sick of talking about it," says Martin. "It's off the table, as far as I'm concerned."

The uncertainty has also led him to explore other ways to assess the quality of graduate education, such as mining existing databases that measure the quantity and quality of scholarly publications. "It'll provide only a subset of the whole picture," Martin admits. "But it's something we can do on our own, inexpensively, and repeat as needed." That's more than the NRC can offer, at least right now.

—JEFFREY MERVIS

Third time, no charm. NRC has put off collecting data for its latest survey of graduate programs.

Published by AAAS

By integrating studies in genomics, neuroscience, and evolution, researchers are beginning to reveal some of the mysteries of animal behavior

A Genomic View of Animal Behavior

Why a dog—or a human for that matter—cuddles up with one individual but growls at another is one of life's great mysteries, one of the myriad quirks of behavior that has fascinated and frustrated scientists for centuries. Here's another: are we hard-wired to tend our young or culturally indoctrinated to have family values?

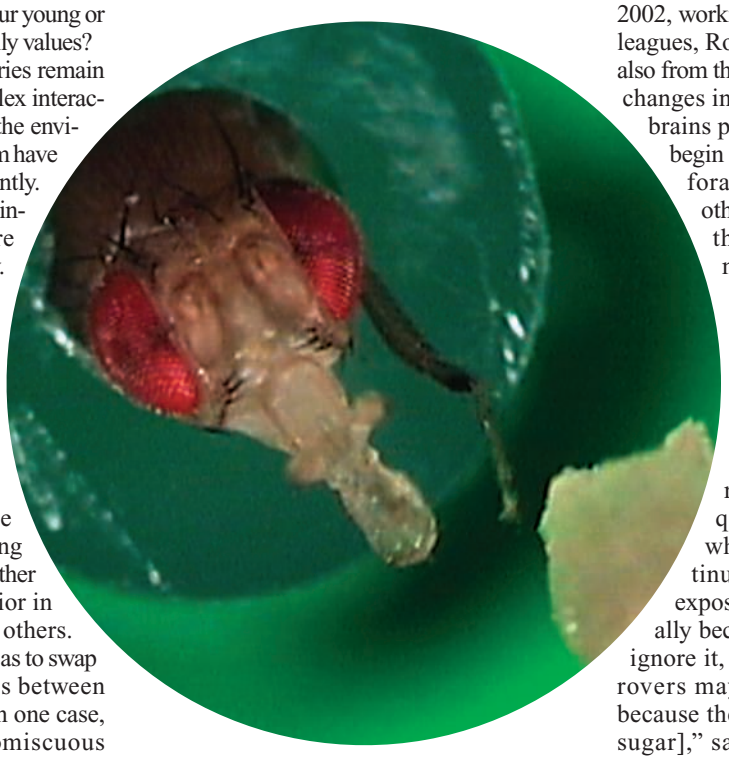
It's no surprise that such mysteries remain unsolved. They are rooted in complex interactions between multiple genes and the environment, and the tools to tackle them have largely been unavailable until recently. But behavioral researchers are beginning to apply techniques that are transforming other areas of biology. They are using microarrays—which can track hundreds or thousands genes at once—to learn, for example, why some honey bees are hive workers and others are foragers, and what makes some male fish wimps and others machos.

They are also comparing the sequenced genomes of the growing menagerie of animals, probing whether genes known to influence behavior in one species play similar roles in others. Investigators have even gone so far as to swap gene-regulating DNA sequences between species with different lifestyles; in one case, they transformed normally promiscuous rodents into faithful partners.

While these comparative approaches are de rigueur for evolutionary biologists, they are something new for many neuroscientists and others who typically study behavior in a single model organism, says Gene Robinson, an entomologist at the University of Illinois, Urbana-Champaign, who is trying to encourage more crosstalk between disciplines. "There is this clear gulf between people who are using modern genetic techniques to study very specific questions and the people who are studying natural diversity," adds Steve Phelps from the University of Florida, Gainesville. But as more behavioral scientists take up the tools of genomics and comparative biology, the payoff may be a deeper understanding of the molecular basis of behavior in animals—even people—and how behaviors originally evolved. The field "is very ripe for a productive synthesis," says Phelps.

Foraging for genes

As gene sequencers turn their attention to deciphering the genomes of dozens of evolutionarily diverse species, a deluge of genome data is beginning to transform some aspects of behavioral science. Instead of just probing the minutiae of how a gene works in one organism, scientists are increasingly investigating how a particular gene operates in multiple species.



Sweet "tooth." A gene that prompts roving in fruit flies also makes them more eager to sip sugar.

Take the story of a wanderlust gene studied by Marla Sokolowski of the University of Toronto, Ontario, Canada. Almost 25 years ago, Sokolowski and her colleagues discovered that a then unidentified gene, which they dubbed *forager* (*for*), controlled how much a fruit fly wandered. One variant of the gene makes a fly a more active forager—a "rover"—while another variant causes a fly to be less active, a "sitter." In 1997, her team finally cloned this gene, which codes for a protein called cGMP-dependent protein kinase (PKG), an important cell-signaling molecule. (*Science*, 8 August 1997, pp. 763, 834) The rover variant turned out to generate higher quantities of the signaling protein.

This gene has recently proved key to feeding behavior in other invertebrates as well. In 2002, working with Sokolowski and her colleagues, Robinson and Yehuda Ben-Shahar, also from the University of Illinois, found that changes in the activity of *for* in honey bee brains prompted hive-bound workers to begin to change roles and start actively foraging for food. That same year, other researchers demonstrated that this gene influenced how likely nematodes were to explore their environment.

In the May-June 2004 issue of *Learning and Memory*, Sokolowski and her colleagues demonstrated that the *PKG* gene affects another behavior — how readily fruit flies respond to sugar. Rover flies are quick to extend their probosis when exposed to sugar and continue to be stimulated by repeated exposure to sugar, while sitters gradually become used to the sweet stuff and ignore it, they reported. "It suggests that rovers may keep on searching for food because they don't [become indifferent to sugar]," says Sokolowski. This constant movement may be an evolutionary advantage for rovers in places where fruits and other foods are scattered.

Given the apparent importance of *for* in the behavior of fruit flies and other species, Sokolowski and Mark Fitzpatrick from the University of Toronto, have now looked across the animal kingdom for the gene and others related to it. They searched public gene databases, and earlier this year, in the February *Journal of Integrative and Comparative Biology*, they reported finding 32 *PKG* genes from 19 species, including green algae, hydra, pufferfish, and humans. The strong sequence conservation of the genes between many species hints that they may play a role in food-related behavior in many organisms. "By studying [*for*] in additional species, we will find out how it modulates foraging behavior in different evolutionary scenarios," says Sokolowski.

CREDIT: RICARDA SCHEINER

Published by AAAS

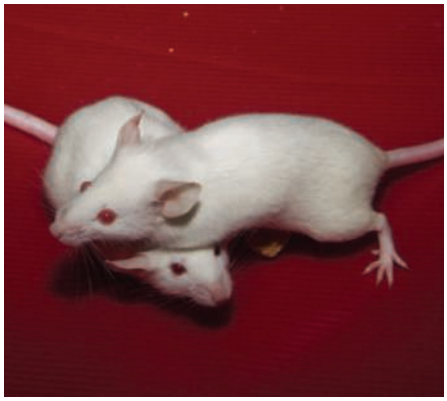
The buzz about microarrays

Comparative genomics is helping researchers pinpoint specific genes involved in some behaviors, but scientists are also using microarrays to cast a broader net. For example, Robinson, behavioral geneticist Charles Whitfield, and their colleagues at the University of Illinois are using these gene expression monitors to study honey bee behavior. They first used microarrays to look at the differences, beyond the *PKG* gene, that distinguish bees that tended the hives from bees that left the hive for pollen (*Science*, 10 October 2003, p. 296). Of the 5500 genes examined, they found 2200 whose brain activity varied between the two types of bees.

Now they have begun to tease out the role of the hive environment in stimulating "nurse" or "forager" genetic regimes—finding genes that help regulate the *PKG* gene's activity. They raised newly emerged bees with no exposure to other bees, then used microarrays to test how certain chemicals known to change bee behavior alter the isolated insects' genetic activity. Last year, Christina Grozinger, now at North Carolina State University in Raleigh, showed that a hormone produced by the queen bee shifted gene expression toward the nurse profile, possibly by suppressing the *for* gene. Ben-Shahar conducted a similar experiment using a hormone that promotes foraging behavior. About half of the genes in the isolated bees shifted in a forager-like direction—and those typically active in hive worker bees turned off.

"We had no genes going in the wrong direction," says Whitfield. Now he and his colleagues are looking at gene expression patterns in bees that either build combs or remove dead bees from a hive. The effort may provide a handle on which genes might promote these construction and undertaker behaviors.

Neurobiologist Hans Hofmann of Harvard University uses microarray technology to probe the behavior of fish. He's investigating the genetic basis for the presence of studs and social outcasts among male cichlids. Some macho males sport bright colors, bully their peers, and court females. Others, the wimps, have small gonads and spend most of their time feeding or swimming in schools with other wimps. In certain circumstances, however, wimps become studs and vice versa, switches that seem to be driven by changing environments.



Close contact. Overly friendly mutant mice helped clarify the genetic pathway involved in reactions to strangers.

CREDITS (TOP TO BOTTOM): Z. HUANG/MICHIGAN STATE UNIVERSITY; E. CHOLERIS/UNIVERSITY OF GUELPH, ONTARIO

Social status. It takes many genes to transform hive workers into foragers.

In the traditional approach, Hofmann would have tried to track individual genes involved in these transformations. Instead, he turned to microarrays and, in less than a year, has identified 100 genes that likely shape the male's social status. Some are genes that Hofmann had expected to be involved, but others, such as a number for ion channels, were surprises. He and his colleagues are now looking more closely at cichlid brains for differences in expression patterns between genes identified in the array studies. "Some of these genes that we decided to follow up, we would not have looked at without this approach," Hofmann notes.

For both Robinson and Hofmann, microarrays have changed the way they investigate genes and behavior. In the pre-genomics era, both chased after candidate genes—those they had reason to suspect were important. But that tunnel vision "doesn't give you a perspective of how many other [genes] are involved," Whitfield explains.

Pathways to behavior

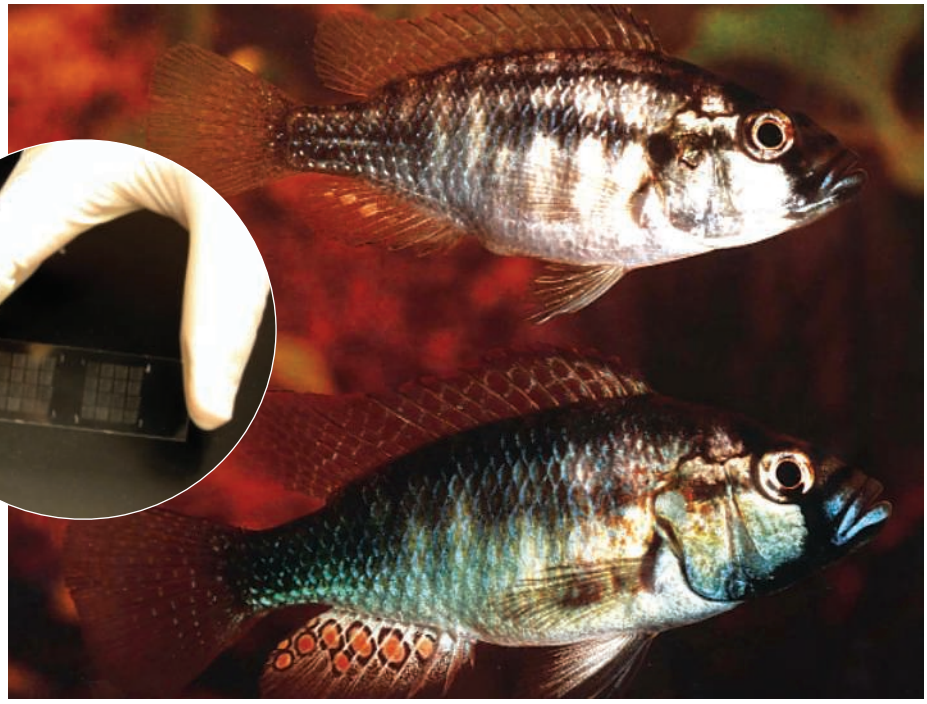
The genetic bounty provided by microarrays poses its own challenges, however. The devices can turn up many genes involved in even a simple behavior, and the molecules those genes encode need to be tied together into a logical pathway. Piecing together that jigsaw puzzle is no easy task.

Elena Choleris from the University of Guelph has taken on that challenge and has worked out the relatively simple pathway underlying one behavioral response in a rodent. She, Martin Kavaliers at the University of Western Ontario, London, Canada, and Don Pfaff from Rockefeller University in New York have shown the genetic interactions necessary for one mouse to recognize another and to react in a friendly or unfriendly manner.

Researchers have known for several years that at least four proteins are involved in this process of social recognition: two estrogen receptors, located in different parts of the brain, and a neuropeptide, oxytocin, and its receptor. Choleris looked at the interplay of these molecules by breeding mutant mice lacking each component. In different groups of mice, she and her team disabled one of the genes encoding the receptors or oxytocin. No matter the genetic defect, the outcome was the same: The mutant mice couldn't tell a familiar mouse from a stranger and were no longer worried about newcomers.

Published by AAAS

From additional experiments, Choleris has deduced some of the protein connections in what she calls a micronetwork, or micronet: One of the estrogen receptors controls oxytocin production in the hypothalamus, while the other receptor works in the amygdala to control the production of oxytocin's receptor. If any component of this micronet is interrupted, the whole pathway breaks down. The micronet exemplifies "how multiple genes act in parallel in an orchestrated manner between different systems and different brain areas," says Choleris. In the wild, a breakdown of this particular micronet and the resulting social recognition deficits could have powerful implications. Choleris and colleagues have recently found that her mutant mice have a diminished ability to sense and stay away from nearby mice carrying parasites, for example.



All-out gene search. Microarrays (inset) are helping to uncover the genes that make some male cichlids more macho (lower fish) than others.

Beyond the gene

Microarrays are powerful tools for spotting genes that underlie different behaviors, but the way those genes are regulated may be just as important as the proteins they produce. Take the case of the prairie vole and the meadow vole.

The prairie vole (*Microtus ochrogaster*) is faithful to its mate; meadow voles (*Microtus pennsylvanicus*) are not. Yet the DNA sequence for vasopressin, the neuropeptide governing this trait, is the same in both species, as is the sequence of the gene for the hormone's receptor protein. There are, however, significant species differences in the number of brain receptors for vasopressin: Prairie voles have a lot more.

Mother's touch. Standoffish mother rats cause chemical changes in DNA bases that make pups timid adults.

In 1999, Larry Young, a neuroscientist at Emory University in Atlanta, Georgia, noticed that a regulatory region, a DNA sequence that sits at the beginning of the receptor gene, was longer in the monogamous species. When he put the prairie vole's vasopressin receptor gene and its regulatory region into mouse embryos, the resulting adult rodents were more faithful than is typical for that particular mouse species. The same has now proved true for meadow voles, he and his colleagues reported in the 17 June *Nature*. When he put the full prairie vole gene, including the regulatory region, for this receptor into meadow voles, males abandoned their promiscuous ways and began acting like faithful prairie voles.

Michael Meaney from McGill University in Montreal, Quebec, has found that a different regulatory region, called a promoter, is pivotal in another social relationship, the one between parents and their offspring. In the early 1990s, he and others had demonstrated that when a mother rat fails to lick and groom her newborn pups, those pups grow up timid and abnormally sensitive to stress.

The key seems to be methylation, a process in which DNA sequences are chemically modified by the addition of methyl groups to cytosine bases. This often suppresses the activity of a gene. Meaney's team discovered that in mice, a mother's behavior alters the typical methylation of the promoter for the gene for the glucorticoid receptor in her offspring. In the brain, this receptor protein helps set off the cascade of gene expression that underlies the stress response.

Before birth, there's no methylation of this gene promoter. But in mice neglected by their mothers, the promoter is methylated shortly after birth, Meaney and his colleagues reported in the 27 June online *Nature Neuroscience*. This increased methylation causes less of the receptor to be produced, creating anxious animals. And because DNA methylation tends to last the life of the animal, it could explain why the pups' personalities don't change as they mature, Meaney notes.

While most behavioral genetics researchers have concentrated on nonhuman species, some are now slowly venturing into the murky waters of human behavior. Meaney's team, for example, is following 200 mothers and their children, looking at the interplay between maternal care and activity in key genes in the offspring. "The extent to which researchers are finding similar patterns" between animals and people is quite promising, notes Stephen Suomi, a psychologist at the National Institute of Child Health and Human Development, Laboratory of Comparative Ethology, Bethesda, Maryland.

These patterns are prompting new research alliances. Genes can represent a common ground, increasing "the links between individuals interested in [neural] mechanisms and the people who are interested in behavior," explains Andrew Bass, a neuroethologist at Cornell University in Ithaca, New York. With this common ground will come a greater understanding of the brain as it relates to behavior, says Pfaff. And that, he adds, "is exciting to the nth degree."

—ELIZABETH PENNISI

CREDITS (TOP TO BOTTOM): R.D. FERNALD AND H.A. HOFMANN; INSET: SUSAN C. P. RENN AND A. HOFMANN; M. MEANEY/MCGILL UNIVERSITY

Source of New Hope Against Malaria is in Short Supply

New drugs based on an old Chinese cure could save countless lives in Africa, if health agencies and companies can find ways to make enough

It seemed like a classic case of bait and switch. In 2004, the World Health Organization (WHO) and the Global Fund for AIDS, Tuberculosis, and Malaria threw their weight behind a radical change in the fight against malaria in Africa. Old, ineffective drugs were to be abandoned in favor of new formulations based on a compound called artemisinin that could finally reduce the staggering death toll. More than 20 African countries have signed on. But the catch is there aren't nearly enough of the new drugs to go around.

Just before Christmas, WHO—which buys the tablets from Novartis for use in African countries—announced that it would deliver only half of the 60 million doses anticipated in 2005, leaving many countries in the cold. "It's a very cruel irony," concedes Allan Schapira of WHO's Roll Back Malaria effort.

Other companies producing the drugs have the same problem as Novartis. Artemisinin is derived from plants grown primarily on Chinese and Vietnamese farms, and they have not kept up with demand. Several plans are afoot to create a new, more stable, and cheaper source. Last month, for instance, the Bill and Melinda Gates Foundation announced a $40 million investment in a strategy to make bacteria churn out a precursor to artemisinin. But such alternatives will take at least 5 years to develop, so the shortages are likely to persist, warns Jean-Marie Kindermans of Médécins sans Frontières in Brussels.

New malaria drugs are badly needed. The parasite *Plasmodium falciparum* has developed resistance to the mainstays, such as chloroquine and sulfadoxine-pyrimethamine. The death toll—more than a million annually—is not declining, despite Roll Back Malaria, an ambitious international campaign launched in 1998 to halve mortality by 2010.

Enter *Artemisia annua* (also known as sweet wormwood or Qinghao), a shrub used for centuries in traditional Chinese medicine. In the 1970's, Chinese researchers discovered that its active ingredient, artemisinin, kills malaria parasites; since then, several chemical derivatives with slightly better properties have been developed. Known by names such as artemether or artesunate, they cure more than 90% of patients within several days, with few side effects observed so far. Best of all, no resistance has been seen yet. To keep it that way, WHO and others recommend that artemisinin compounds always be used with a second drug in a so-called Artemisinin-based Combination Therapy, or ACT.

Widely used in Asia, the introduction of ACTs in Africa has lagged. Countries have been reluctant to make the switch because, at about $2.40 per treatment course, ACTs are 10-20 times more expensive than existing drugs. The Global Fund has also dragged its feet, some allege, by funding the purchase of older, cheaper drugs for too long. Things began to change when an expert group published a scathing letter in *The Lancet* in January 2004, accusing the Global Fund and WHO of "medical malpractice." Both organizations denied the claims, explaining that they supported ACTs but that change took time. Both also concede that the ensuing debate spurred them to redouble their efforts.

Fields of gold. Extracts of *Artemisia annua* (bottom) provide powerful new malaria drugs, but farms have not met demand for the shrub.

But companies are reluctant to produce the drugs, as are farmers to grow *Artemisia*, without guarantees that they'll sell—and that's the problem. The Global Fund does not have nearly enough money to fund the drugs' introduction across Africa. Donor countries like the U.S. and the U.K. appear reluctant to spend aid money on market guarantees for big pharma, says Schapira, because it could be seen as lining shareholders' pockets; at an emergency session at WHO just before Christmas, no donors made any commitments.

WHO's hope is that growing demand will eventually create a stable artemisinin supply at low prices. *Artemisia* farms are now springing up in India, and WHO is supporting experiments to grow the plants in east Africa.

The Gates Foundation is banking on a less fickle supply route. Over the past 10 years, chemical engineer Jay Keasling and colleagues at the University of California, Berkeley, have spliced nine genes into *Escherichia coli* bacteria to make them produce terpenoids, a class of molecules that includes artemisinin. With a few genes borrowed from *Artemisia*, they should be able to produce an artemisinin precursor, Keasling says.

On 13 December, the foundation announced a $42.6 million grant to the Institute for OneWorld Health in San Francisco—which bills itself as the world's first non-profit pharmaceutical company—to help Keasling finish the engineering. Then a biotech startup will optimize the process for producing artemisinin—"tons and tons of it," says OneWorld Health president Victoria Hale—about 5 years from now. Her assumption is that pharmaceutical companies will package OneWorld's artemisinin derivates into ACT tablets and sell them at well under a dollar per treatment.

There's another alternative. Jonathan Vennerstrom and colleagues at the University of Nebraska, Omaha have synthesized a compound called OZ277 (or simply OZ) that, like artemisinin, has a peroxide bridge shielded by large chemical rings. The compound has been tested as an antimalarial in vitro and in animals, and it looks even better than the real thing, Vennerstrom and colleagues reported in *Nature* in August. Ranbaxy, an Indian pharmaceutical company, is developing it further; a phase 1 safety trial has just been completed.

Ideally, 4 or 5 years from now, OZ will result in new drug combinations that have the power of current ACTs but cost less than a dollar per treatment, says Chris Hentschel, chief executive of the Medicines for Malaria Venture (MMV), a non-profit based in Geneva that supports its development. Still, Hentschel is trying to temper his optimism. Drugs can always fail during testing, and even ACTs may eventually lose their efficacy, like almost every malaria drug before. That's why, despite the new hope, MMV has its pipeline well-stocked with unrelated candidates.

—MARTIN ENSERINK

CREDITS (TOP TO BOTTOM): HOLLEY PHARMACEUTICALS; SCOTT BAUER/USDA

Published by AAAS

Oldest Civilization in the Americas Revealed

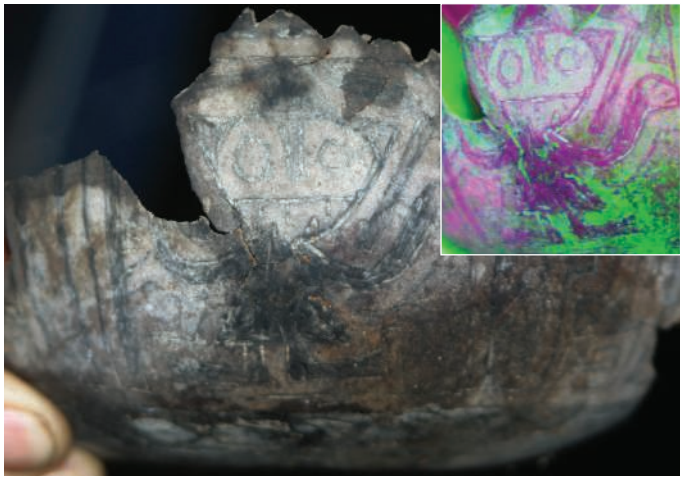
Almost 5000 years ago, ancient Peruvians built monumental temples and pyramids in dry valleys near the coast, showing that urban society in the Americas is as old as the most ancient civilizations of the Old World

BARRANCA, PERU—A few miles northeast of this small fishing town, the Pan-American Highway cuts through a set of low, nondescript hummocks in the narrow Pativilca River valley. If they were so inclined, the truckers thundering along the road could spot on the hillocks the telltale signs of archaeological activity—vertical-sided cuts into the earth surrounded by graduate students with trowels, brushes, tweezers, plastic bags, and digital cameras.

The Pativilca, about 130 miles north of Lima, is one of four adjacent river valleys in the central Peruvian seacoast known collectively as the Norte Chico, or Little North (see map, p. 35). Pinched between rain shadows caused by the high Andes and the frigid Humboldt Current offshore, this is one of the driest places on earth; rainfall averages 5 cm a year or less. Because of the exceptional aridity, ancient remains are preserved with startling perfection. Yet the same aridity long caused archaeologists to ignore the Norte Chico, because the region lacks the potential for the full-scale agriculture thought to be necessary for the development of complex societies.

Then in the 1990s, groundbreaking research directed by archaeologist Ruth Shady Solis of the Universidad Nacional Mayor de San Marcos established that such societies had existed in the Norte Chico in the third millennium B.C.E., the same time that the Pharaohs were building their pyramids (*Science*, 27 April 2001, p. 723). And in the 23 December issue of *Nature*—in what archaeologist Daniel H. Sandweiss of the University of Maine at Orono describes as "truly significant" work—archaeologists Jonathan Haas of the Field Museum in Chicago and Winifred Creamer and graduate student Alvaro Ruiz of Northern Illinois University in DeKalb reported the startling scope of the Norte Chico ruins, which include "more than 20 separate residential centers with monumental architecture," and are one of the world's biggest early urban complexes. The ruins are dominated by large, pyramid-like structures, presumably temples, which faced sunken, semicircular plazas—an architectural pattern common in later Andean societies. The new work includes 95 radiocarbon dates that confirm the great antiquity of this culture, which emerged about 2900 B.C.E. and survived until about 1800 B.C.E.

The concentration of cities in the Norte Chico is so early and so extensive, the archaeologists believe, that coastal Peru must be added to the short list of humankind's cradles of civilization, which includes Mesopotamia, Egypt, China, and India. Yet the Peruvian coast, as Shady has argued, is in some ways strikingly unlike the others. She points out that most of the Eurasian centers "interchanged goods and adaptive experiences," whereas the Norte Chico "not only developed in isolation from those [societies], but also from Mesoamerica, the other center of civilization in the Americas, which developed at least 1500 years later." The result, according to Haas, is that the Norte Chico provides a laboratory in which to observe "that most puzzling phenomenon, the invention of the state." The people of this ancient, isolated society, says Haas, "had no models, no influences, nobody to copy. The state evolved here purely for intrinsic reasons."



Gourd lord. This piece of gourd reveals a figure (shown in false color, inset) carved about 2250 B.C.E. in the Norte Chico region.

CREDIT: COURTESY OF THE FIELD MUSEUM

Cities without farms

Although the Norte Chico mounds were flagged as possible ruins as far back as 1905, researchers never excavated them because, according to Ruiz, "they didn't have any valuable gold or ceramic objects, which is what people used to look for." The first full-scale excavation took place in 1941, when Gordon Willey and John M. Corbett of Harvard discovered a single multiroomed building at Aspero, a salt marsh at the mouth of the Supe River. Puzzled by what seemed to be an isolated structure, the team took 13 years to publish their data.

Willey and Corbett also noted a half-dozen odd "knolls, or hillocks," which the two men described as "natural eminences of sand." Thirty years later, in the 1970s, Willey returned to Aspero with archaeologist Michael E. Moseley, now at the University of Florida at Gainesville. They quickly established that the site actually covered 15 ha and that the natural knolls were, in truth, "temple-type platform mounds." It was "an excellent, if embarrassing, example," Willey later wrote, "of not being able to find what you are not looking for." When carbon dating revealed that the site was very old, Moseley says, "it became obvious that Aspero was something big and important."

It was also a conundrum. All complex Eurasian societies developed in association with large river valleys, which offered the abundant fertile land necessary for agriculture. And social scientists have long believed that the organization of labor necessary for agriculture was the wellspring of civilization. Aspero, on a little river that coursed through a desert, had almost no farmland. "We asked, 'How could it sustain itself?'" Moseley says. "They weren't growing anything there, or almost anything."

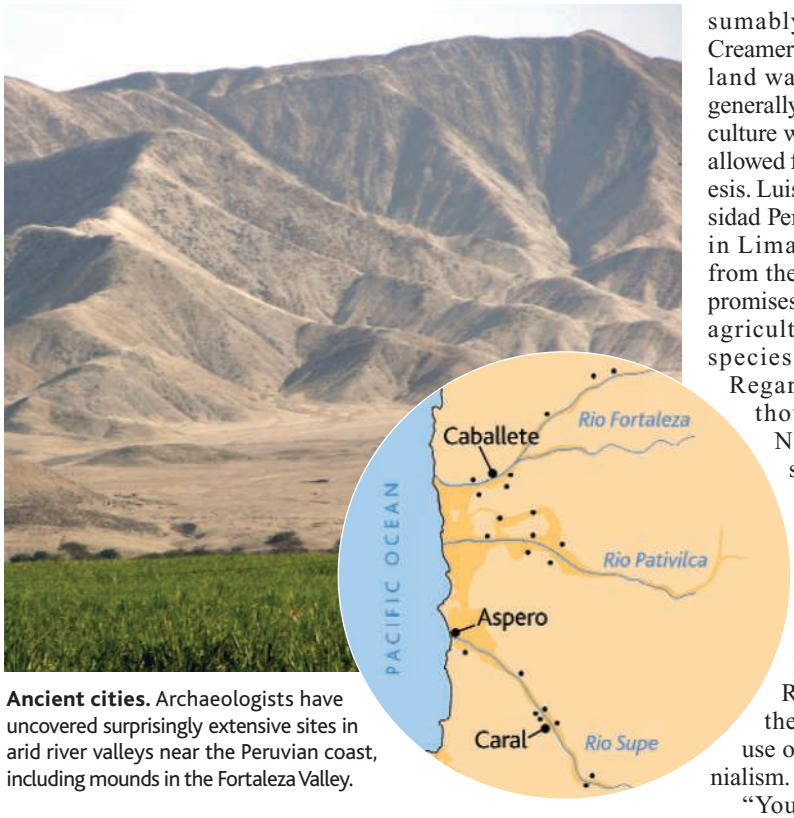
The question prompted Moseley in 1975 to draw together earlier work by Peruvian and other researchers into what has been called the MFAC hypothesis: the maritime foundations of Andean civilization. He proposed that there was little agriculture around Aspero because it was a center of fishing, and that the later, highland Peruvian cultures, including the mighty Inca, all had their origins not in the mountains but in the great fishery of the Humboldt Current, still one of the world's largest. Bone analyses show that late-Pleistocene coastal foragers "got 90% of their protein from the sea—anchovies, sardines, shellfish, and so on," says archaeologist Susan deFrance of the University of Florida, Gainesville (*Science*, 18 September 1998, pp. 1830, 1834). "Later sites like Aspero are

Published by AAAS

just full of [marine] fish bones and show almost no evidence of food crops." The MFAC hypothesis, she says, boils down to the belief "that these huge numbers of anchovy bones are telling you something."

Despite its explanatory power, the hypothesis had to be modified when Shady began work at Caral, almost 23 kilometers upriver from Aspero. One of 18 sites with monumental and domestic architecture found by Shady's team, Caral covered 60 ha and was, in Shady's view, a true city—a central location that provided goods and services for the surrounding area and was socially differentiated, with lower-class barrios in the periphery and elite residences with painted masonry walls in the center. Dating to about 2800 B.C.E., Shady says, Caral was dominated by a pyramid bigger than a football field at the base and more than seven stories high, overlooking a plaza bordered by smaller monumental structures. The big buildings suggested a large resident population, but again there were plenty of anchovy bones and little evidence of subsistence agriculture. The agricultural remains were mainly of cotton, used for fishnets, and the tropical tree fruits guayaba and pacae. All were the products of irrigation. At the Norte Chico, the Andes foothills jut close to the coast, creating the sort of swiftly dropping rivers that are easiest to divert into fields.

To Moseley, the abundance of fish bones at Caral suggested that the ample protein on the coast allowed people to go inland and build irrigation networks to produce the cotton needed to expand fishing production. Caral thus lived in a symbiotic relationship with Aspero, exchanging food for cotton.



Ancient cities. Archaeologists have uncovered surprisingly extensive sites in arid river valleys near the Peruvian coast, including mounds in the Fortaleza Valley.

CREDIT: PROYECTO ARQUEOLOGICO NORTE CHICO; INSET: ADAPTED FROM J. SEAGARD/THE FIELD MUSEUM

The making of a state

The central structures in Norte Chico cities were constructed in what Haas believes to have been sudden bursts of as few as two or three generations. The buildings were made largely by stacking, like so many bricks, mesh bags filled with stones. So perfect is the preservation that the Peruvian-American team can remove 4,000-year-old "bricks" from the pyramids almost intact, the cane mesh around them still in place. (Along with food remains, the mesh provided many of the samples used for carbon dating.) But the impressive size of the monuments is not matched by a rich material culture; the Norte Chico society existed before ceramics.

According to Creamer, Haas, and Ruiz, the sheer scale of the inland sites raises a major challenge to the MFAC hypothesis. "The great bulk of the population lived inland in these cities," Creamer says. "If there were 20 cities inland and one on the coast, and many of the inland cities are bigger than the coastal city, the center of the society was inland."

But defenders of the MFAC hypothesis remain convinced that the coastal areas were of primary import. "What may be important," says deFrance, is not the scope of the society "but where it emerged from and the food supply. You can't eat cotton."

Whether maritime or inland cities developed first, it seems clear that each depended on the other, and Haas says that this interdependency has major implications. "If I look beyond Aspero at this time, what I see is a bunch of fishing sites all up and down the Peruvian coast. All of them have cotton, but they are on the coast where they can't really grow it. And then you have one big gorilla inland—a concentration of inland sites that are eating anchovies but can't obtain them themselves. It's a big puzzle until you put them together. ... I believe we are getting the first glimpses of what may be the growth of one of the world's first large states, or something like it."

In archaeological theory, societies are often depicted as moving from the kin-based hierarchy of the band to the more abstract authority of the state in order to organize the defense of some scarce resource. In the Norte Chico, the scarce resource was presumably arable land. Haas, Creamer, and Ruiz think that the land was more valuable than generally believed, and that agriculture was more important than allowed for in the MFAC hypothesis. Luis Huaman of the Universidad Peruana Cayetano Heredia in Lima is examining pollen from the Norte Chico sites and promises that "we will see when agriculture came in and what species were grown there."

Regardless of the results, though, the cities of the Norte Chico were not sited strategically and did not have defensive walls; no evidence of warfare, such as burned buildings or mutilated corpses, has been found. Instead, Haas, Creamer, and Ruiz suggest, the basis of the rulers' power was the use of ideology and ceremonialism.

"You have lots of feasting and drinking at these sites," Haas says. "I have the beginning of evidence that there are the remains of feasts directly incorporated into the monuments, the food remains themselves and the hearths from cooking all built into it." Building and maintaining the pyramids—so unlike anything else for thousands of miles—was the focus of communal spiritual exaltation, he suggests. A possible focus for the religion is the curious figure Creamer found incised on a gourd. Dated to 2250 B.C.E., it resembles in many ways later Peruvian deities, including the Inca god Wiraqocha, suggesting that the Norte Chico may have founded a religious tradition that existed for almost 4000 years.

Despite their excitement about the new work, MFAC backers see no reason yet to give up the belief that, as Sandweiss puts it, "the incredibly rich ocean off this incredibly impoverished coast was the critical factor." Only the upper third of Aspero has been excavated, notes deFrance, and its emergence has never been properly dated. If coastal Aspero, though much smaller than the inland cities, is also much older, the MFAC hypothesis might be supported. With Moseley, Shady's team is hoping to resolve the debate by digging to the bottom of Aspero next summer. Meanwhile, Haas, Creamer, and Ruiz have bought a house in Barranca for a laboratory and barracks. They plan to work in the area for years to come. "You're going to be hearing a lot more about the Norte Chico," Ruiz promises.

—CHARLES C. MANN

Published by AAAS

Is Tobacco Research Turning Over a New Leaf?

Scientists developing reduced-harm tobacco products increasingly rely on tobacco industry funding, but some universities and grant organizations want to forbid it

A 65-year-old man sitting at a small table in a lab at Duke University Medical Center in Durham, North Carolina, asks for a cigarette, his twelfth in less than eight hours. A researcher is happy to oblige. As the man lights up, a swarm of technicians buzz around him, drawing blood from a catheter in his arm, making him exhale into a sensor, and administering cognitive tests.

The experiment, led by neuroscientist Jed Rose, focuses on the volunteer's response to a cigarette called Quest, made from tobacco that's been genetically engineered to contain less nicotine. Rose directs the university's Center for Nicotine and Smoking Cessation Research, dedicated to helping smokers kick the habit. He sees the Quest study as an important step in the center's mission because it indicates that smokers of this new product inhale less deeply than smokers of an earlier "reduced-harm" product—the low-tar cigarette—and may therefore be able to decrease their dependence on tobacco. But the work is controversial. Quest's maker, the Vector Tobacco Company of Research Triangle Park, North Carolina, paid for the study, and tobacco giant Philip Morris funds the center.

Since the late 1990's the tobacco industry has provided university researchers with millions of dollars to help develop a new class of reduced-harm products—including modified cigarettes like Quest, tobacco lozenges, and nicotine inhalation devices—ostensibly to reduce the hazards of smoking. Advocates say the industry has turned over a new leaf and is now serious about improving the safety of its products. But critics, who cite the industry's efforts to manipulate science over the past 50 years, see nothing but the same old smoke and mirrors.

Anti-smoking activists tried to stop tobacco's research juggernaut more than a decade ago—and won some battles. But industry funding is flourishing, igniting a debate on some campuses over whether universities should ban tobacco money and whether grant organizations should deny funding to individuals or schools that take this money—as Britain's Wellcome Trust already does and the American Cancer Society is about to do.

It's not a simple issue, says Ken Warner, a public health expert at the University of Michigan, Ann Arbor, and president of the Society for Research on Nicotine and Tobacco. He concedes that the tobacco industry was guilty of misconduct in the past but worries about restricting research. "How do you avoid infringing on academic freedom, and what sort of slippery slope do you create by denying grants on moral grounds?" he asks. "There is a real need for reduced-harm research. The question is, given their history, do we let the tobacco companies fund it?"

Burning issue. University of Nebraska's Stephen Rennard says bans on tobacco industry funding violate academic freedom.

Moral dilemma

Duke University's Rose thinks the tobacco industry's new focus on harm reduction may usher in a healthier era of tobacco-sponsored research. This research is "high quality, innovative, and unique," he says, and "very different from the abuses of the past." He adds, "None of the companies that fund our studies have made any attempt to bias our work."

Rose, a co-inventor of the nicotine patch, argues that vilifying the industry won't help the millions of smokers who are trying to quit. "The real enemy is the death and disease smokers suffer," he says. "If we can use tobacco money to help people lead healthier lives, why shouldn't we?"

Stephen Rennard, a pulmonary physician at the University of Nebraska Medical Center in Omaha who also receives tobacco industry support, agrees. "I approach this from a public health perspective," he says. "People are going to continue to smoke, and we need to make them as safe as we can. The tobacco industry needs university research to develop a safer product."

One of Rennard's projects, funded by RJ Reynolds, evaluated Eclipse—a standard-looking cigarette manufactured by the company that heats rather than burns tobacco, theoretically producing less harmful smoke. Rennard later used Philip Morris money to determine how much smoke the average cigarette user is exposed to. The findings may help the company design a cigarette that reduces the levels of inhaled smoke.

Still, Rennard says that taking industry money required a lot of soul searching. "But in the end I realized that this research *should* be funded by tobacco companies. NIH resources should not be used to improve cigarettes. It would be like the government subsidizing the development of a better laundry detergent."

"It's trendy to beat up on the tobacco industry," Rennard adds. "It's simplistic, and it doesn't help the people who need to be helped. If we delay this research because of concerns about tobacco funding, it could be years before these potentially life-saving products make it to market. That would be the real tragedy."

Smoky past

Others think academic researchers should just say no to tobacco money. Simon Chapman, editor of the journal *Tobacco Control* and a professor of public health at the University of Sydney in Australia, says that despite their new efforts to support harm reduction studies, the tobacco companies have little interest in public health. "They fund this research to buy respectability and ward off litigation," he says. Some worry that reduced-harm products are just a ploy to keep smokers addicted. Chapman says that scientists need only look at current examples of tobacco company malfeasance—from targeting youth smokers in Myanmar to using athletes to promote cigarettes in China—to see that the companies haven't changed their ways.

For many critics of mixing tobacco money with university research, the industry's history speaks for itself. For example, as the link between smoking and disease

CREDIT: RON ERTI/UNIVERSITY OF NEBRASKA MEDICAL CENTER

Published by AAAS

became clearer in the early 1950's, the world's largest tobacco companies established the Tobacco Industry Research Committee (TIRC)—later the Council for Tobacco Research (CTR)—to fund research into the health effects of smoking. But its main goal, internal company documents now reveal, was to obfuscate risks, and few of the studies it funded addressed the hazards of cigarettes (*Science*, 26 April 1996, p. 488).

"During the four decades they operated, TIRC and CTR never came to the conclusion that smoking causes cancer," says Michael Cummings, the director of the Tobacco Control Program at the Roswell Park Cancer Institute in Buffalo, New York. "These organizations were more about public relations than science." The industry agreed to shut down CTR in 1998 as part of an agreement—known as the Masters Settlement—that also awarded 46 U.S. states $206 billion in compensation for the cost of treating smoking-related illnesses.

But CTR wasn't the only problem. Government prosecutors have charged that the companies frequently killed their own research when it came to unfavorable conclusions, funded biased studies designed to undermine reports critical of smoking, and used the names of respected scientists and institutions to bolster their public image. The industry also lost credibility with its previous attempts at harm reduction when it touted low-tar and filtered cigarettes introduced in the 1950's and '60's as "safer," says Chapman, while suppressing evidence that smokers drew harder on these cigarettes, thereby *increasing* their uptake of carcinogens. These charges are being revisited in an ongoing federal racketeering case—the largest civil lawsuit in American history—alleging a 50-year conspiracy by the tobacco industry to mislead the public about the dangers of smoking. For its part, the industry argues that it has reformed; Philip Morris spokesperson Bill Phelps says his company believes that investing in research is the best way to address the health risks associated with smoking.

Richard Hurt, the director of the Nicotine Dependence Center at the Mayo Clinic in Rochester, Minnesota, says researchers considering industry money should remember the toll extracted by tobacco use—4.9 million deaths per year worldwide, according to World Health Organization estimates. "For anyone interested in public health, taking this money is a clear conflict of interest," he says.

Harm reducer? RJ Reynold's Eclipse heats rather than burns tobacco, theoretically producing less harmful smoke.

Academic freedom

While scientists debate the merits of taking tobacco money, other authorities may take the decision out of their hands. Over the past decade, a number of institutions—including the Harvard School of Public Health and the University of Glasgow—have prohibited their researchers from applying for tobacco industry grants. In addition, organizations such as Cancer Research U.K. and the Wellcome Trust will no longer fund researchers who take tobacco money. The American Cancer Society, one of the largest private funders of cancer research, plans to adopt a similar policy this month.

Ohio State University, Columbus, found itself in the eye of the storm in 2003 when Philip Morris offered a medical school researcher a $590,000 grant at the same time a state foundation offered a nursing school researcher a $540,000 grant. Because the terms of the state grant would have prohibited all other university researchers from taking tobacco money, the school could not accept both. "There was a very heated debate among the faculty," says Tom Rosol, the university's senior associate vice president for research, who ultimately made the decision to take the Philip Morris grant. "It came down to the issue of academic freedom," he says. "We didn't want to accept a grant that would have placed restrictions on our investigators." Rosol's decision sparked a backlash, and several departments, including the Comprehensive Cancer Center and the School of Public Health, enacted tobacco funding bans, barring researchers from taking tobacco money in the future.

No sale. University of Sydney's Simon Chapman says the tobacco industry wants to buy researchers' credibility.

A resolution approved by the University of California's (UC) Academic Senate this summer would have the opposite effect. Stating that "no special encumbrances should be placed on a faculty member's ability to solicit or accept awards based on the source of funds," the proposal would forbid any institutions within the UC system from banning tobacco funding. In a letter endorsing the resolution, UC president Robert Dynes describes such bans as "a violation of the faculty's academic freedom."

Not everyone buys the academic freedom argument. "The university should be a role model," says Joanna Cohen, an expert on university tobacco policies at the University of Toronto. "Academic freedom should not override its ethical responsibilities."

Nor does the American Legacy Foundation, a Washington, D.C., tobacco education and funding organization established by the Masters Settlement, have any qualms about denying grants to institutions that take tobacco money. "We don't see this as an academic freedom issue," says Ellen Vargyas, the foundation's general council. "The tobacco industry has a bad history, and this is our way of encouraging institutions not to take their money."

The University of Nebraska's Rennard, who made himself ineligible for state money by accepting tobacco industry funds, finds these policies and the university bans deeply flawed. "Political positions should not determine scientific agendas," he says. "If we restrict research on moral grounds, should we ban grant money from pharmaceutical companies or industries that pollute the environment? Where do you draw the line?"

As public funding gets tighter, more universities may find themselves confronting this question. The tobacco industry is poised to fill the financial void, but continued charges of company malfeasance will increase the pressure on schools to shun this money. At the end of the day, institutions will have to decide whether to overlook the source of this funding, or take the moral high road and watch it go up in smoke.

–DAVID GRIMM

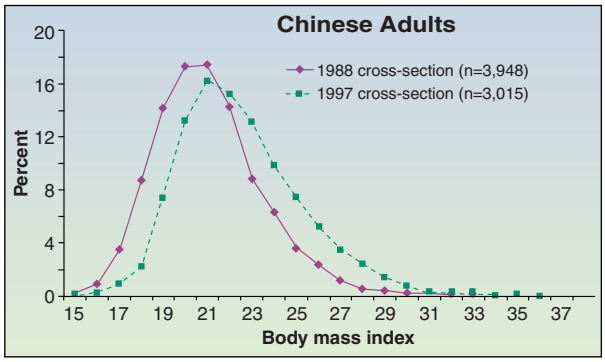
CREDITS (TOP TO BOTTOM): RJ REYNOLDS/ASSOCIATED PRESS; COURTESY OF SIMON CHAPMAN

Published by AAAS

Random Samples

Edited by Constance Holden

Obesity in the East



You don't have to be wealthy any more to be sedentary and overweight. As this chart shows, the Chinese are gaining according to measurements of body mass index. They're also experiencing "a marked decline in physical activity," according to a new report from the U.S. National Academies. The report, "Growing up Global: The Changing Transitions to Adulthood in Developing Countries," says the proportion of Chinese children who are overweight increased from 6% to 8% in a 7-year period.

Toughening Up Brain Cells

Scientists have demonstrated a direct link between a very low calorie diet and resistance to Parkinson-like symptoms in rhesus monkeys—the first time this has been observed in primates.

Researchers at the National Institute on Aging in Baltimore and the University of Kentucky in Lexington put seven male rhesus monkeys, out of a group of 13, on a bare-bones diet for 6 months. They then injected a neurotoxin into one side of the brain in all 13 monkeys to produce Parkinson-like symptoms. Compared to their free-eating colleagues, the dieting monkeys maintained significantly higher levels of locomotor activity after the injections.

Post mortem analysis of movement-related brain areas showed that the dieters also had slightly higher levels of dopamine, a neurotransmitter that is depleted in Parkinson's patients, and about three times the levels of GDNF, a nerve growth factor.

This suggests that caloric restriction may protect brain cells by turning up the production of growth factors, and suggests that a long-term calorie-controlled diet might reduce the risk of Parkinson's disease in humans, the scientists report in a paper published online last month in the *Proceedings of the National Academy of Sciences*. Neuroscientist Ole Isacson at Massachusetts General Hospital in Belmont says the results confirm that calorie restriction somehow makes the brain cells "tougher." But he cautions that the molecular mechanisms behind that effect remain to be identified.

Giant Eagle had Modest Origins

Using DNA extracted from 2000 year-old bird bones, scientists have discovered that an extinct giant eagle from New Zealand was descended from an Australian bird one tenth its size.

Haast's eagle was the biggest the world has ever seen. It weighed about 12 kilograms, had a wingspan of up to 3 meters, and had talons as big as tiger's claws and strong enough to puncture the pelvic bones of the huge flightless birds that it dined on.

New Zealand paleobiologist Richard Holdaway had surmised that the bird was descended from the 4.5kg Australian wedge-tailed eagle. But an analysis of mitochondrial DNA shows a closer tie with eagles of the South Pacific that only weigh about one kilogram, Holdaway's team reports in the 4 January issue of the *Public Library of Science Biology*. They say the New Zealand eagle underwent an exceptionally large and rapid size increase once it settled on the islands, probably benefitting from being at the top of the food pyramid with no predators in prehistoric New Zealand, which saw its first human settlers only 700 years ago.

Asians have low cancer rates; Latinos getting more melanomas.

California Cancers

A racial and ethnic breakdown of the nearly 2 million Californians diagnosed with cancer from 1988 through 2001 reveals dramatic group differences in the disease. In a report released last month by the California Cancer Registry, epidemiologists Myles Cockburn and Dennis Deapen of the University of Southern California's Cancer Surveillance Program tracked 23 types of cancer in 9 major ethnic groups. Among unexplained findings, says Cockburn, is that South Asians have the lowest cancer incidence of all groups. Among women, Koreans had the lowest breast cancer rate. Rates of skin cancer are rising in Latino populations. Other findings confirm previous data—for example, blacks have the highest prostate cancer mortality, with rates 10 times as high as in Asians.

"It is well known that there are androgen receptor differences between races" that may contribute to the prostate cancer gap, says geneticist Joel Buxbaum of Scripps Research Institute in La Jolla. But screening programs are probably responsible for the decline in cervical cancer among immigrant Vietnamese women and the lowered breast cancer mortality rate in all but Filipino women. Observers say the report should serve as a goldmine for other researchers. "The level of detail is what makes it unique," says biostatistician Brenda Edwards from the National Cancer Institute in Bethesda, Maryland.

Published by AAAS

PEOPLE

Edited by Yudhijit Bhattacharjee

2005 People to Watch

Tectonic shift. Only at the Massachusetts Institute of Technology (MIT) would a group called the Logarhythms serenade the school's new president. But their warbling is music to the ears of neuroscientist Susan Hockfield, who has pledged to sing the praises of MIT's faculty and students "to kids in our own nation and around the world" as part of a broader effort to "reinvigorate science and technology education" in the United States.

Hockfield arrives 5 years after MIT publicly admitted it was discriminating against its women scientists. Researchers and administrators around the country will be watching to see how Hockfield, who took over last month from Chuck Vest, will change the playing field for women in science.

Enough tarrying. As head of the European Commission's directorate general for research, economist Achilleas Mitsos has a mandate from his political bosses to build the $6-billion International Thermonuclear Experimental Reactor (ITER) in France. But he'd rather have all six ITER partners, including Japan, on board than split the collaboration. In addition to those seemingly endless negotiations, Mitsos will also be pushing to double the budget of the European Union's next 5-year research program, which starts in 2006, and launch a new basic research agency, the European Research Council.

Up in the air. The space science spotlight in 2005 will be on Al Diaz, the new head of a reorganized science office within the National Aeronautics and Space Administration (NASA). A dyed-in-the-wool technocrat who joined NASA in 1964 and most recently ran Goddard Space Flight Center in Greenbelt, Maryland, Diaz will have the thankless task of slicing and dicing research programs to fit into a 2005 budget that, despite increasing, falls far short of what the agency needs. With Sean O'Keefe headed out the door, Diaz will also need to sell that science to a new NASA administrator.

The Peacemaker. Look for nuclear watchdog Mohamed ElBaradei to take on an even higher profile this year as the director general of the International Atomic Energy Agency (IAEA). In addition to reining in ambitious nuclear programs in Iran and North Korea and watching developments in Brazil and South Korea, the Egyptian-born lawyer will also be trying to secure a third term despite blatant U.S. efforts to oust him by searching for a viable replacement.

Finding coherence. The burden of defending Europe against infectious diseases rests on the shoulders of Hungarian epidemiologist Zsuzsanna Jakab. Nominated last month to lead the new European Centre for Disease Prevention and Control (ECDC) in Stockholm, Jakab must figure out how a small outfit with no labs can help agencies and labs in 25 countries battle SARS, influenza, and other threats to public health. If confirmed, Jakab would be the first agency head from one of the 10 countries that joined the European Union in 2004.

National Shrinking Foundation. Arden Bement could be the first director in the 55-year history of the National Science Foundation (NSF) to preside over consecutive years of declining budgets. Bement, who started a 6-year term in late November after spending most of the year as acting NSF director, is already coping with a 2% cut in 2005 imposed by Congress. And the Bush Administration has told NSF to prepare for a 4% cut in the president's 2006 budget request next month. Bement is also trying to find heads for three of NSF's seven research directorates, one of which has been vacant since last March.

EVENT WATCH

Look for...

- the Department of Energy (DOE) to pick a new contractor to run Lawrence Berkeley National Laboratory. The University of California's contract to manage the lab expires 31 January, but it will stay in charge until DOE announces its decision later in the year.
- the National Institute of Allergy and Infectious Diseases to announce the team of academics that will run the new Center for HIV/AIDS Vaccine Immunology, a virtual lab that will receive at least $300 million over the next 7 years. The center will support a research program aimed at addressing immunological challenges that stand in the way of an effective HIV vaccine.
- the launch of the California Institute of Regenerative Medicine, the centerpiece of a 10-year, $3 billion plan for stem cell research approved by the state's voters last year. The institute plans to award its first grants by May.
- NASA's Space Shuttle to return to orbit to continue building the International Space Station.

Published by AAAS

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 6 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

Ethics of Rationing the Flu Vaccine

IT IS GOOD TO SEE THAT THE CENTERS FOR Disease Control are seeking ethical guidance about the rationing of flu shots this year ("Ethicists to guide rationing of flu vaccine," J. Couzin, News of the Week, 5 Nov., p. 960). They should also be seeking ways to reduce the scientific uncertainties that make this ethical question difficult. For example, how effective is the influenza vaccine in different age groups, and in preventing disease, mortality, and risk of transmitting to others? How much benefit do hospital patients or nursing home residents receive through reduced risk of transmission if the staff of these institutions are vaccinated, compared with the benefits of being vaccinated themselves? Surely the answers to these questions are important inputs to the ethical calculus, and existing data are not adequate to answer them.

Attempts to solve this short-term ethical problem should not obscure the larger failures that led to it. It has been common knowledge for years that the influenza vaccine supply was fragile at best, and no serious effort has been made to ensure a safe and plentiful supply. This failure also increases the risk of delays and and limited supplies of the new vaccine that will be required for a pandemic. Other pandemic preparedness activities, including surveillance for new strains in Asia, are seriously underfunded.

Big failures of policy and politics have led to ethical dilemmas that would not otherwise have existed. While offering their guidance for the short term, perhaps the ethical advisory board can also highlight the ethical and other benefits of taking actions to minimize the likelihood of rationing in the future.

MARC LIPSITCH

Department of Epidemiology, Harvard School of Public Health, 677 Huntingdon Avenue, Cambridge, MA 02115, USA.

Jellyfish Blooms in the Yangtze Estuary

JELLYFISH BLOOMS IN ESTUARIES WORLDWIDE (*1*) can have substantial effects on plankton communities and fish populations because jellyfish are consumers of zooplankton and ichthyoplankton. Jellyfish populations normally fluctuate regularly, causing periodic blooms. Perhaps the most damaging type of jellyfish blooms have been caused by nonindigenous species. Although there are little available data, there seems to be general agreement that human activities are having measurable effects on the oceans and coastal habitats and that jellyfish blooms are responses to these changes (*2*).

***Sanderia malayensis* Goette**

CREDIT: W. XIAN *ET AL.*

The Three Gorges Dam across the Yangtze River is the world's largest hydroelectric dam and will form a huge reservoir of 1080 km^2 (*3*). The common jellyfish in the Yangtze Estuary was an edible species, *Rhopilema esculenta* Kishinouye, which suffers from overfishing. It has gradually been replaced by *Cyanea capillata* Linnaeus, which has bloomed in summer and sometimes until autumn since 1997 (*4*). The Three Gorges reservoir began to store water on 1 June 2003. At the water storage stage, the discharge of water and sediments into estuary is greatly reduced, making the saltwater intrusion appear earlier and the duration of intrusion longer. The increase of water temperature and salinity, the high level of nutrients, and the abundance of zooplankton have stimulated the expansion of *Cyanea capillata*, from a prevalence of 0.41% in 1998 to 85.47% of the total samplings for fisheries in November 2003 (*5*).

In May 2004, there was a bloom of the jellyfish *Sanderia malayensis* Goette (see the figure). *Sanderia malayensis* is mainly distributed in the West Indian Ocean and the Arabian Sea and has been recorded in China only in the South Sea (*6*). This was the first time it bloomed in the Yangtze estuary, and it now makes up 98.44% of total catches, clogging the mesh of the trawl nets. The surface of the estuary was almost completely covered in some places. Currents and ships may be responsible for the appearance of *Sanderia malayensis*, and the conditions caused by reduced discharge (and particularly the increased temperature) may favor the growth of this species.

The Yangtze River mouth and surrounding sea is one of China's major fisheries, but these resources have been decreasing. Jellyfish compete directly with fish for food. Removal of predator fish throughout the world's oceans by commercial fishing efforts (*7*) allows the jellyfish populations to expand (*8*), with concomitant blooms. May and November are the two fish-spawning seasons in the Yangtze Estuary. Jellyfish blooms in these periods can drive the fish away, and jellyfish also feed on the fish eggs and larvae, bringing destructive aftereffects to the Yangtze estuarine fisheries resources in the current and coming years.

WEIWEI XIAN,[1] BIN KANG,[1,2] RUIYU LIU[1]

[1]Key Laboratory of Marine Ecology and Environmental Sciences, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, People's Republic of China. E-mail: wwxian@ms.qdio.ac.cn. [2]Fishery College, Ocean University of China, Qingdao 266003, People's Republic of China.

References and Notes

1. J. E. Purcell *et al.*, *Mar. Ecol. Prog. Ser.* **180**, 187 (1999).
2. C. E. Mills, *Hydrobiologia* **451**, 55 (2001).
3. P. Xie, *Science* **302**, 1149 (2003).
4. X. Zhong *et al.*, *Modern Fish. Inf.* **19** (no. 3), 15 (2004).
5. State Environmental Protection Administration, Bulletin on the Ecological and Environmental Monitoring Results of the Three Gorges Project (1999, 2001, 2003, 2004).
6. H. Hong *et al.*, *J. Xiamen Fish. Coll.* **7** (no. 2), 7 (1985).
7. D. Pauly *et al.*, *Science* **279**, 860 (1998).
8. C. E. Mills, F. Sommer, *Mar. Biol.* **122**, 279 (1995).

9. We acknowledge support from the Three Gorges Project Construction Committee in the item "The Three Gorges Project and the Estuarine Ecology and Environment" (SX2001 – 018). We also thank D. S. Mclusky for his constructive suggestions and review on this paper.

Does Aneuploidy or Mutation Start Cancer?

IN HIS ARTICLE "DISEASE BACKS CANCER ORIGIN theory," David Grimm makes a case for the theory that mutation causes cancer via aneuploidy (the state of having an abnormal number of chromosomes) (News of the Week, 15 Oct., p. 389). The "starting gun" is said to

be "mutations in a gene involved in ensuring proper chromosome number," i.e., *BUB1B*, based on a recent study of the heritable mosaic variegated aneuploidy syndrome by Rahman *et al.* (*1*). According to Rahman, the resulting "aneuploidy is a cause, not an effect of cancer," a conclusion also supported by Lengauer and Wang (*2*).

However, the probability is very low that mutation of a specific gene is the "starting gun" of carcinogenesis through aneuploidization in all those without inherited defects in mitosis genes, because carcinogens or spontaneous accidents induce aneuploidy much more effectively than specific mutations.

First, it is much more likely that aneuploidization is initiated when a carcinogen, e.g., an x-ray, strikes at a random point along a chromosome than at one specific gene. Because humans carry about 35,000 genes and have a large complement of noncoding DNA, the target of mutating a specific aneuploidy gene is vanishingly small compared with the target of direct aneuploidization through losing, fragmenting, or rearranging a chromosome. Moreover, the discrepancy between the two targets is even bigger, if one considers that the spindle apparatus is also a target of aneuploidization (*3*). For example, "cytoplasmic" radiation of the spindle apparatus has recently been shown to induce aneuploidy (*4*). Second, about 50% of the known carcinogens are not even mutagens (e.g., asbestos), and are thus not able to induce aneuploidy or cancer through mutations (*3*, *5*).

In view of this, my colleagues and I have recently proposed that carcinogenesis is initiated by a random aneuploidy, which is generated either by a carcinogen or spontaneously (*3*, *6*, *7*). Because aneuploidy unbalances numerous teams of proteins, which segregate, synthesize, and repair chromosomes, it destabilizes the numbers and structures of chromosomes. Owing to this inherent instability, aneuploidy catalyzes a chain reaction of chromosomal evolutions. Aneuploidy is thus a source of chromosomal variations from which, in classical Darwinian terms, selection would encourage the emergence of new cell "species" with neoplastic phenotypes and karyotypes. Nevertheless, cancer is neither a fast nor a necessary consequence of aneuploidization. This follows, because the probability to generate by random chromosome reassortments a new cell species, which outcompetes a normal cell with a 3-billion-year history of evolution, is very low.

PETER DUESBERG

Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA.

References

1. S. Hanks *et al.*, *Nature Genet.* **36**, 1159 (2004).
2. C. Lengauer, Z. Wang, *Nature Genet.* **36**, 1144 (2004).
3. P. Duesberg, A. Fabarius, R. Hehlmann, *IUBMB Life* **56**, 65 (2004).
4. J. B. Little, *Carcinogenesis* **21**, 397 (2000).
5. W. Lijinsky, *Environ. Mol. Mutagen* **14** (suppl. 16), 78 (1989).
6. P. Duesberg, D. Rasnick, *Cell Motil. Cytoskel.* **47**, 81 (2000).
7. P. Duesberg, R. Li, *Cell Cycle* **2**, 202 (2003).

Orphan Enzymes?

IN SPITE OF THE WEALTH OF KNOWLEDGE gathered during the past 60 years about enzymes and their activities, well-curated databases such as UniProt (*1*, *2*) lack sequences for more than 40% of well-characterized enzymes, despite the fact that these databases contain 1,215,979 amino acid sequences.

Enzyme functions are classified by the international Enzyme Commission (EC) (*3*) according to a four-digit system, the EC numbers. As of April 2004, each of 3820 EC numbers corresponded to an enzymatic activity that has been unambiguously defined and approved by the Nomenclature Committee (*3*). Of these 3820 enzymes, 42.5% do not have any sequence available in UniProt (release 1.7 of 13 April 2004) (48, 46, 37, 38, 39, and 28%, respectively, for classes 1 to 6). Considering the presence of many enzymes in numerous species and the availability of complete sequences of some 200 organisms, these numbers were surprising to us.

The distribution of the 1625 enzyme functions without amino acid sequences was independent of the age of their discovery: The relative percentage of sequenceless EC numbers was higher after 1978, i.e., after the beginning of the gene/genome sequencing era, than before (50.5% and 35.4%, respectively). Even in the recent period of intense genomic sequencing (2001–04), this remains true; of the 274 newly described enzyme activities, 105 (38.3%) are sequenceless.

In many cases, the molecules putatively responsible for enzyme functions have been identified but their amino acid sequence has not yet been deciphered. For example, a malate oxidase activity (first published in 1959, created in 1961 as EC 1.1.3.3) has been discovered in *Micrococcus lysodeikticus* and further characterized in several other bacteria. These include entirely sequenced organisms such as Mycobacteria and *Escherichia coli* in which the mode of action of the candidate protein, including regulation of its expression and its allosteric properties, were described in detail in 1979 (*4*).

Another case is D-mannitol oxidase activity (first published in 1986, created in 2001 as EC 1.1.3.40), which occurs in snails within a specialized tubular organelle, the mannosome (*5*). The mannosome proteins have been purified, and the candidate protein responsible for the EC 1.1.3.40 activity has been identified in Western blots (*6*) but is not yet sequenced.

Around half (51.6%) of the sequenceless EC numbers have been observed in a single organism (or group of closely related organisms). This may reflect specific metabolic requirements of a peculiar lifestyle, as in the case of the mannosome-related oxidase. Such peculiarities combined with the absence of reliable genetic tools could explain the difficulties of getting cognate sequences.

A few of the sequenceless EC numbers could correspond to annotation errors, as in a case we have described (*7*). The genes encoding putrescine carbamoyltransferases (EC 2.1.3.6) have been erroneously annotated in several completely sequenced organisms as ornithine carbamoyltransferases (EC 2.1.3.3). From a forgotten partial sequence, we were able to reassign the correct sequences to EC 2.1.3.6, thought to be devoid of sequence. Such an approach could be more widely applied by scanning the so-called "bibliome" (*8*) or by reconsidering old lab books, doctoral theses, and similar unpublished material.

OLIVIER LESPINET AND BERNARD LABEDAN

Institut de Génétique et Microbiologie, CNRS UMR 8621, Université Paris Sud, Bâtiment 409, 91405 Orsay Cedex, France E-mail: olivier.lespinet@igmors.u-psud.fr, bernard.labedan@igmors.upsud.fr

References

1. R. Apweiler *et al*. *Nucleic Acids Res.* **32**, D115 (2004).
2. See www.expasy.org/sprot/.
3. Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB), www.chem.qmul.ac.uk/iubmb/enzyme/index.html.
4. S. Narindrasorasak *et al*., *J. Biol. Chem*. **254**, 1540 (1979).
5. J. E. Vorhaben *et al*., *Int. J. Biochem.* **18**, 337 (1986).
6. T. Knigge *et al*., *Comp. Biochem. Physiol. C Toxicol. Pharmacol*. **131**, 259 (2002).
7. D. G. Naumoff *et al*., *BMC Genom*. **5**, 52 (2004).
8. L. Grivell, *EMBO Rep*. **3**, 200 (2002).

Amyloidosis and Protein Folding

IN HIS PERSPECTIVE "IN THE FOOTSTEPS OF alchemists" (28 May, p. 1259), C. M. Dobson concisely summarizes the critical role of protein misfolding and aggregation in the development of amyloidosis. He also details several therapeutic approaches to this disease based on inhibiting or reversing aggregation, or removing aggregated proteins. But aggregation does not kill or damage cells; aggregated proteins do, and how they kill remains uncertain. A growing body of evidence suggests these aggregates damage cells by forming ion-permeable channels in cellular membranes (*1*–*6*). Nearly all amyloid peptides studied so far have been found to form channels (*7*). These nonspecific "leakage" pathways would drain cellular

energy stores, inhibit neuronal signaling, disrupt Ca^{++} homeostasis (*8*), and trigger apoptosis. The role of membrane disruption by islet amyloid polypeptide in the death of pancreatic beta cells is now well established (*9*), and a similar role for other amyloid peptides such as Abeta, prion protein (PrP106-126), serum amyloid A, beta-2-microglobulin, and nonamyloid aggregates such as huntingtin (polyglutamine) has been proposed (*10*). The propensity of amyloid peptides to form ion channels is no doubt a direct result of their physical chemical properties and the suitability of beta-sheets for forming pore structures (*11*).

This evidence suggests at least two other potential therapeutic approaches to amyloid disease: (i) Membrane "stabilizing" agents could be developed to prevent the insertion of amyloid channels into lipid membranes (*12*). (ii) Channel-blocking compounds could be selected using known amyloid channels inserted into lipid bilayers. Channel blockers such as Zn^{++} have been shown to protect fibroblasts from Abeta toxicity (*4*).

It has also been proposed that at least one amyloid protein, serum amyloid A (SAA), might have a role in host defense, perhaps killing invading microbes by channel formation (*6*). If this turns out to be true for other amyloids, we may have to once again rethink our notions about the biological function of amyloid and alter our therapeutic goals accordingly.

BRUCE L. KAGAN
Department of Psychiatry, UCLA School of Medicine, 760 Westwood Plaza, 67-468 NPI, Los Angeles, CA 90024–1759, USA.

References
1. N. Arispe, E. Rojas, H. B. Pollard, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 567 (1993).
2. T. Mirzabekov, M. Lin, B. L. Kagan, *J. Biol. Chem.* **271**, 1988 (1996).
3. M. C. Lin, T. Mirzabekov, B. L. Kagan, *J. Biol. Chem.* **272**, 44 (1997).
4. R. Bhatia, H. Lin, R. Lal, *FASEB J.* **14**, 1233 (2000).
5. Y. Hirakura, B. L. Kagan, *Amyloid* **8**, 94 (2001).
6. Y. Hirakura, I. Carreras, J. D. Sipe, B. L. Kagan, *Amyloid* **9**, 13 (2002).
7. J. I. Kourie, A. L. Culverson, P. V. Farrelly, C. L. Henry, K. N. Laohachai, *Cell Biochem. Biophys.* **36**, 191 (2002).
8. F. M. LaFerla, *Nature Rev. Neurosci.* **3**, 862 (2002).
9. J. Janson, R. H. Ashley, D. Harrison, S. McIntyre, P. C. Butler, *Diabetes* **48**, 491 (1999).
10. B. L. Kagan, R. Azimov, R. Azimova, *J. Membr. Biol.*, in press.
11. C. Petosa, R. J. Collier, K. R. Klimpel, S. H. Leppla, R. C. Liddington, *Nature* **385**, 833 (1997).
12. N. Arispe, M. Doh, *FASEB J.* **16**, 1526 (2002).

Response

KAGAN'S LETTER HIGHLIGHTS A TOPIC OF GREAT debate in the amyloid field, namely, the specific mechanism through which misfolded and aggregated proteins can cause disease. One of the most interesting aspects of all the "misfolding" diseases from my point of view is that, despite their many differences, the underlying origins of these diseases could be remarkably similar (*1*). Moreover, these origins appear to stem from the intrinsic physicochemical properties of polypeptide chains and the way that proteins have co-evolved with the environments in which they function.

The "generic" model for amyloid formation and disease (*2*) is completely consistent with, although by no means dependent on, the hypothesis that ion channels could play a major role in the way misfolded proteins damage cells (*3*). Indeed, the effects on cells in culture of misfolded proteins having no connection with disease have been found to be closely similar to the effects of aggregates of the peptides and proteins that are associated with conditions such as Alzheimer's disease (*4*). Moreover, and in accord with Kagan's comments, they are clearly linked to calcium homeostasis (*5*). In addition, it is evident that the potential toxicity of any misfolded or aggregated protein is generally held at bay by the natural defenses of

cells and organisms, including molecular chaperones and targeted degradation mechanisms (*6*). Thus, approaches to therapy that focus on this aspect of aggregation behavior are an inherent part of a "generic" picture of protein misfolding diseases.

The additional point made by Kagan in his Letter, namely, that at least one protein known to form amyloid aggregates in vivo could have a role in host defence or other "normal" biological processes, is also an extremely important one, and one for which there is increasing evidence (*7*). Another example, in addition to the one mentioned by Kagan, that is also of great interest involves the remarkable properties of the calcium-binding protein α-lactalbumin. This protein is present in milk, and it appears that under some circumstances, it can misfold and give rise to species that have the ability to kill cancer cells in a selective manner. Not only are such species likely to be naturally protective for infants who are breastfed, but they are now in clinical trials as potential cancer therapeutics (*8*). Biology often has the power to surprise us, and these examples of cases where potentially fatal forms of proteins can have positive benefit under some circumstances are likely to be the tip of a large iceberg.

CHRISTOPHER M. DOBSON
Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK.

References
1. C. M. Dobson, *Nature* **426**, 884 (2003).
2. C. M. Dobson, *Trends Biochem. Sci.* **9**, 329 (1999).
3. M. Stefani, C. M. Dobson, *J. Mol. Med.* **82**, 678 (2003).
4. M. Bucciantini *et al.*, *Nature* **416**, 507 (2002).
5. M. Bucciantini *et al.*, *J. Biol. Chem.* **279**, 31374 (2004).
6. F. U. Hartl, M. Heyer-Hartl, *Science* **295**, 1852 (2002).
7. J. Kelly, W. E. Balch, *J. Cell Biol.* **161**, 461 (2003).
8. L. Gustaffson *et al.*, *N. Engl. J. Med.* **350**, 2663 (2004).

CORRECTIONS AND CLARIFICATIONS

Reports: "A mammalian H^+ channel generated through alternative splicing of the NADPH oxidase homolog *NOH-1*" by B. Bánfi *et al.* (7 Jan. 2000, p. 138). This Report described three mRNA products of the *NOH-1* gene (now called NADPH Oxidase 1, *NOX1*): NOH-1L (NOX1α), NOH-1Lv (NOX1β), and NOH-1S (NOX1γ). NOX1α and β are encoded by 13 and 12 exons of *NOX1*, respectively, and are similar in length. NOX1γ, however, is much shorter and is encoded by 6 exons: exons 1 to 4, part of exon 5, and exon 14. T. Leto and M. Geiszt questioned whether NOX1γ is a genuine splice variant. Indeed, although NOX1α and β mRNA were readily detected by the Northern blot technique, the authors could detect NOX1γ mRNA only by RT-PCR or by RNAse protection. The sequence of the complete 3′ untranslated region of NOX1α by Shu *et al.* (GenBank AF127763) indicates that "exon 14" is not a separate exon but is located at the very end of exon 13. Analysis of the secondary structure of NOX1α mRNA with mfold software [M. Zuker, *Nucleic Acids Res.* **31**, 3406 (2003); D. H. Mathews, J. Sabina, M. Zuker, D. H. Turner, *J. Mol. Biol.* **288**, 911 (1999)] revealed that nucleotides at position 658 to 672 could form a stable double helix with nucleotides 2449 to 2462, with only one unpaired nucleotide. This helix would bring a CCCAUCC motif at 675 to 681 very close to the same CCCAUCC motif at 2460 to 2466, with the potential to generate a loop that would allow transcriptional slippage [Y. J. Zhang, H. Y. Pan, S. J. Gao, *Biotechniques* **31**, 1286 (2001)]. To investigate this possibility, the authors used different reverse transcriptases to generate cDNA from mRNA of the human colon epithelial cell line, CaCo-2, which they had suggested contained NOX1α and γ transcripts. PCR measurements with primers to exons 4 and 13 demonstrated that a Moloney murine leukemia virus (MuMLV) reverse transcriptase (Multiscribe), but not two template slippage-resistant reverse transcriptases (Omniscript and Thermoscript) generated NOX1γ cDNA (fig. S1) (see Supporting Online Material available at www.sciencemag.org/cgi/content/full/307/ 5706/44/DC1). Thus, the authors conclude that NOX1γ is not a genuine isoform, but an artifact most likely due to a stable loop formation of the NOX1 mRNA. Although this observation does not invalidate the authors' conclusion that transfection of this short form of NOX1 confers H^+ current to cells, the short form is a truncated form of NOX1 and not a naturally occurring splice variant. The authors apologize for any confusion that this mischaracterization has caused.

HUMAN ECOLOGY

Learning from the Past to Change Our Future

Tim Flannery

Jared Diamond's acclaimed *Guns, Germs, and Steel* (*1*) tells the story of humanity's rise from the hunter-gatherer societies of 13,000 years ago to the organized states in which most of us live today. *Collapse* is a perfect sequel, for it examines the fate that may be in store for our societies in the next few decades. While he planned the book, Diamond at first thought that it would deal only with human impacts on the environment. Instead, what has emerged is arguably the most incisive study of senescing human civilizations ever written.

Five factors guide Diamond's analysis: cumulative environmental damage, climate change, hostile neighbors, friendly trade partners, and society's response to all of these. The book entails a broad-ranging and complex analysis that demands mastery of diverse disciplines—from ecology to climatology, sociology, politics, and history. This is the sort of thing at which Diamond excels, yet *Collapse* would be nowhere near as powerful a work without his acute understanding of the human condition—particularly the motivations, limits of perception, methods of organization, and mental flexibility that are the common lot of humanity.

Diamond begins his analysis on familiar territory: the dairy farms of Montana, where he worked as a farmhand while a student. He has known the landscape and the people of this spectacular region for half a century and during that time has seen a dramatic transformation. Foremost among the changes he chronicles is the "conquest" of the Montana pioneers by wealthy out-of-staters, who in the absence of effective planning laws (some Montana counties even lack building codes and zoning laws) have built dude ranches, housing estates, and industrial developments at their whim. The result is that many former ranchers are now landless menial workers who labor in the estates of the wealthy new settlers—an outcome that has engendered considerable soul-searching. As Diamond puts it, "Montanans are beginning to realize that two of their most cherished attitudes are in direct opposition: their pro-individual-rights anti-government-regulation attitude, and their pride in their quality of life." This conflict of values is a key theme to which he returns again and again throughout the book.

The bulk of *Collapse* is taken up by considerations of societies that have failed (including Easter Island, the Classic Maya, and the Greenland Norse) and of societies such as the Tikopians, Tokugawa-era Japanese, and Icelanders, which have survived against the odds. Diamond places great store on the capacity of environmental conditions to shape society, which some may see as a bias toward environmental determinism. However, his fifth factor—how people react to environmental challenges—puts paid to such ideas. The Greenland Norse provide an example of particular relevance to our contemporary world. Inhabitants of a new and different land, they clung to a Christian, European lifestyle that ultimately doomed them to extinction. "It was out of the question to invest less in churches, to imitate or intermarry with the Inuit, and thereby to face an eternity in Hell just in order to survive another winter on Earth," Diamond says of the decisions that doomed them.

Collapse
How Societies Choose to Fail or Succeed
by Jared Diamond
Viking Press, New York, 2005. 591 pp. $29.95, C$44. ISBN 0-670-03337-5.

Classic collapse. At Maya sites such as Tikal, the population, number of monuments and buildings, and number of Long Count calendar dates plummeted during the 9th century A.D.

Diamond frames the Rwandan genocide as a contemporary example of a society in collapse. It was not, he argues, simply a racially motivated massacre, for the murders also occurred in areas where just one ethnic group (Hutu or Tutsi) was present. The real tension was over land. With median farm size declining from 0.89 acres in 1988 to 0.72 acres in 1992 and with inequality increasing, large sections of Rwandan society were driven to desperation in a classic Malthusian tragedy.

The last chapters of *Collapse* are devoted to the contemporary developed world. The perilous state of the Australian environment gives Diamond reason to suspect that Australia may be the first developed state to collapse under environmental pressures. This may seem absurd to many affluent Australians, but Diamond demonstrates convincingly that societies typically collapse when at the height of their dynamism and affluence, because that is precisely when resource demand is greatest. One thing, however, is on Australia's side: its people are forging a new relationship with their land and in the process discarding cultural baggage such as sheep grazing, which came from England and in the past was a source of great wealth. This Diamond sees as a great positive because "the values to which people cling most stubbornly under inappropriate conditions are those values that were previously the source of their greatest triumphs."

In the final chapter, Diamond reflects on his own society, the United States. Many of his friends make great sacrifices so that their children can attend the best (and most expensive) schools, yet they barely give a thought to the environment in which their children will mature. The situation has now become so dire, Diamond believes, that huge changes to our societies will probably occur within the next few decades. Yet he is a cautious optimist who sees in growing environmental awareness and new technology reason to hope that we can triumph over adversity.

Diamond's book will doubtless spawn many sceptics and naysayers, including the likes of the CEO of one American mining company who believes that "God will soon arrive on Earth, hence if we can just postpone land reclamation for another 5 or 10 years it will then be irrelevant anyway." Yet the fact that one of the world's most original thinkers has chosen to pen this mammoth work when his career is at its apogee is itself a persuasive argument that *Collapse* must be taken seriously. It is probably the most important book you will ever read.

The reviewer is at the South Australian Museum, North Terrace, Adelaide, South Australia 5000, Australia. E-mail: flannery.tim@saugov.sa.gov.au

References
1. J. M. Diamond, *Guns, Germs, and Steel: The Fates of Human Societies* (Norton, New York, 1987).

10.1126/science.1105671

Published by AAAS

ASTRONOMY

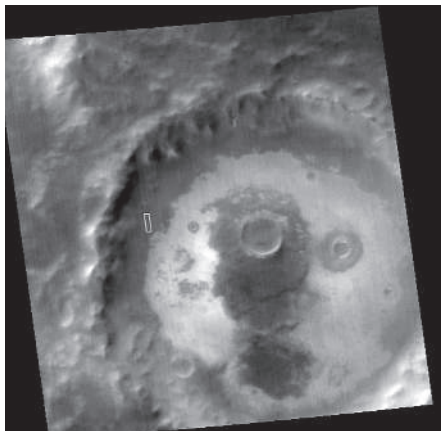
A Field with a Life of Its Own

Jeffrey L. Bada

Today, it seems nearly everyone is an astrobiologist. A decade ago, I knew essentially none. Why this sudden obsession with a field that encompasses so many diverse areas in both the physical and life sciences? So far, life has not been found to exist away from Earth, although the surge in interest in astrobiology suggests there is intense optimism within at least parts of the science community that this singularity will change in the future. But scientific curiosity alone likely cannot explain the explosive growth of astrobiology. After reading *The Living Universe: NASA and the Development of Astrobiology*, I came to the conclusion that one of the field's attractions was money—and lots of it.

Steven Dick (NASA, Washington, D.C.) and James Strick (Franklin and Marshall College, Pennsylvania) are historians of science, and thus in the book the growth of astrobiology as a distinct scientific discipline is extensively detailed and referenced. Soon after NASA was formed in 1958, the agency began funding research in exobiology, the older relative of astrobiology. The initiation of this effort was likely influenced by Joshua Lederberg, whose 1960 paper defining exobiology as the study of life beyond Earth set out the justifications for this research topic (*1*). For an agency initially formed to advance the space program of the United States, NASA's funding of exobiology research was a bold step. Some of these early exobiology grants were indeed substantial; between 1959 and 1964, for example, there were awards of approximately $2 million for a space science building at the University of California at Berkeley, $1.7 million to Sidney Fox, and $1.5 million to Lederberg. In addition to funding research at academic institutions, NASA also supported exobiology studies at some of its centers, notably the Exobiology Division at the Ames Research Center (ARC) at Moffett Field. Through the 1960s and 1970s, NASA continued to sponsor a robust exobiology program, which was a major funding source for the development of the life-detection instruments sent to Mars on Viking. In the post-Viking era, however, the exobiology program shifted toward basic research with generally smaller grants. The change left places like ARC with the problem of how to sustain a robust exobiology program. Time was ripe for a new research initiative.

The Living Universe
NASA and the Development of Astrobiology
by Steven J. Dick and James E. Strick
Rutgers University Press, Piscataway, NJ, 2004. 288 pp. $49.95. ISBN 0-8135-3447-X.



Commemorative crater. Tikhov's efforts in astrobiology have been acknowledged in the naming of this impact crater on Mars.

In the mid-1990s, NASA administrators Wesley Huntress and Daniel Goldin envisioned astrobiology as a means of integrating biological sciences into the space exploration program while also revitalizing places such as ARC and providing a solid funding base for academic research. In the spring of 1995, Goldin officially designated ARC as NASA's center for astrobiology. Then, summer 1996 brought the announcement of supposed evidence for life in the martian meteorite ALH84001 (*2*). Among scientists and the general public alike, this claim generated intense interest in—as well as controversy about—the possibilities of life beyond Earth. All of a sudden, astrobiology was the hottest topic around.

Capitalizing on this enthusiasm, in the fall of 1997 NASA announced the first round of competition for its Astrobiology Institute nodes. Envisioned as a "virtual" institute, the Institute was to be housed at ARC with the nodes spread among NASA centers, universities, and independent research institutions. The scientific community raced to get a piece of the action, and today the Institute comprises 16 funded nodes with five-year budgets of between about $5 million and $12 million. Several international partners have joined the Institute, demonstrating the global reach of the field. New journals dealing exclusively with astrobiology have appeared, and there are yearly conferences or sessions at national and international meetings dealing with the subject. The spring 2004 Astrobiology Science Conference at ARC attracted some 700 registrants. The field has indeed exploded.

The Living Universe provides rich documentation of the history of NASA's involvement in exobiology and astrobiology, and in general I found the account to be readable and informative. However, at times the chapters seem disconnected, suggesting they were written individually and then pasted together (e.g., several people and places are introduced multiple times). Readers who know nothing about the field and its participants may find some parts hard to follow. Rightly or wrongly, some individuals are glamorized and championed as major players in the field, others are demonized merely for holding strong opinions, and some key contributors are barely mentioned, if at all. For example, I found the discussion that tries to justify the substantial early funding awarded to Sidney Fox rather odd because the impact of his research today is generally considered minimal. Moreover, reading the book leaves the impression that NASA invented exobiology and astrobiology, which is incorrect. The term astrobiology was first used in 1941 by Laurence Lafleur, who defined the field as "the consideration of life in the universe elsewhere than on earth" (*3*)—a definition that remains a central part of the Astrobiology Institute's roadmap. And in 1953, the Russian astrophysicist Gavriil Tikhov published *Astrobiology*, a book focused on the spectra of plants and his attempts to detect vegetation on Mars (*4*). Researchers from Europe, Japan, and elsewhere have made significant contributions to the field, but Dick and Strick do not mention them. There are also factual errors: for example, Stanley Miller did not help design an instrument for Mariner 4.

The next couple of decades will bring a concerted effort by both NASA and the European Space Agency to ascertain whether life ever existed, or possibly still exists, on Mars or Europa. Astrobiology has a big stake in these efforts. Finding evidence for life on another body in our solar system would give the field the justification it requires in order to remain an active, well-funded area of research. If finding evidence for life continues to be elusive, then as George Gaylord Simpson once noted, astrobiology will remain an area of study without a known subject.

References
1. J. Lederberg, *Science* **132**, 398 (1960).
2. D. S. McKay *et al.*, *Science* **272**, 924 (1996).
3. L. Lafleur, *Astron. Soc. Pac. Leafl.* **143**, 333 (1941).
4. G. A. Tikhov, *Astrobiologiya* (Molodaya gvardia, Moscow, 1953).

10.1126/science.1106678

The reviewer is at the Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093–0212, USA. E-mail: jbada@ucsd.edu

CREDIT: NASA/JPL/MALIN SPACE SCIENCE SYSTEMS

Protector of the Seeds: Seminal Reflections from Southern Africa

Patricia Berjak

Despite their marked geographical and cultural diversity, the peoples of Africa are bound together by concerns about food security and the vagaries of rainfall across the continent's extensive terrain, much of which is arid or semi-arid. This makes the scientific study of seeds and their storage an imperative. I became convinced of this scientific mandate even as a graduate student at the University of Natal in Durban in the late 1960s where, under the guidance of Trevor Villiers, I metamorphosed from an animal-oriented biochemist into a seed-focused cell biologist.

To most people, a seed is a dry structure that can be maintained in a desiccated condition in a state of suspended animation until provided with water and other conditions that will promote germination. These traits define "orthodox" seed behavior. Maize (corn), which produces orthodox seeds, is the staple crop of much of Africa, yet it is ill-suited to the drought-prone conditions that prevail in many regions, where it is cultivated in preference to the native cereal, sorghum.

Annual production of maize is important not only for food security, but also in providing seeds for planting in following seasons. Unfortunately, the crop is frequently jeopardized by droughts. The threat to the crop is exacerbated by seed storage under warm, high relative humidity conditions that can drain seeds of their vigor and viability, while encouraging fungal growth in the seeds. My doctoral work on maize seeds aimed to characterize the course of rapid deterioration that inevitably occurs under these poor seed-storage conditions. I concentrated on the root cap of the seed embryo. After germination, the integrity of this structure is essential to protect the tip of the root as it grows through the sharp, abrasive soil.

This year-long essay series celebrates 125 years of *Science* by inviting researchers from around the world to provide a regional view of the scientific enterprise.

Radical Fungi

There were several exciting outcomes from those investigations. The first was the original microscopical characterization in plant cells of lysosomal vacuoles—fluid-filled vesicles collectively containing enzymes capable of breaking down all other intracellular constituents. A second discovery was that cells that form the root cap self-destruct by autolysis in the final phase of their developmental program (a process called apoptosis, or programmed cell death) and are sloughed at the cap surface. The work also showed that the events involved in apoptosis are accelerated when seeds are poorly stored and that intracellular membranes are the primary loci of degeneration.

Membranes are pivotal for compartmentalizing intracellular functions. They also provide the selective barrier between the cell and its surroundings. Membrane breakdown is a key factor in cell debilitation and death. For seeds, that translates into a loss of viability. Then, as now, the generation of free radicals within the cells of dry seeds in storage is considered to be a major cause of deterioration of membranes and other cellular structures.

On the premise that membrane damage is caused by free-radical activity, Norman Pammenter, my husband and major research collaborator, and I had an inspiring discussion with a Hungarian animal physiologist, K. Molnár, about his work on the efficacy of cathodic protection in extending the lifespan of mice.

Consequently, we stored maize seeds under deteriorative conditions, but in a static electric field. The results, published 30 years ago in this journal, showed that the application of cathodic protection had a dramatic effect in extending seed lifespan. That outcome could be attributable to quenching of free radicals. With hindsight, however, another interpretation is also possible: The efficacy of the treatment resulted from its adverse effects on fungi within the seeds.

With the help of two graduate students, David Mycock and Michelle McLean, my laboratory became active in seed fungus research. The fungi in question are xerotolerant—they survive the dry conditions within stored orthodox seeds. They also produce mycotoxins, which include some

CREDIT: DENISE BERJAK

Patricia Berjak
South Africa

Patricia Berjak, a professor at the University of KwaZulu-Natal in Durban, South Africa, has studied seeds and seed storage for over 30 years. A biochemist initially, she metamorphosed early in her career into a cell biologist with a passionate focus on seeds. She works with her husband and scientific colleague, Norman Pammenter, and many of her graduate students have moved on to distinguished careers of their own. She is a recipient of the 2004 Distinguished Woman Scientist Award, administered by the South African Department of Science and Technology, and the Silver Medal of the South African Association of Botanists. She was recently inducted as a Fellow of the Third World Academy of Sciences. In addition to her scientific pursuits, she enjoys high-performance cars, light-aircraft aviation, and ballroom dancing.

All essays appearing in this series can be found online at www.sciencemag.org/sciext/globalvoices/

Published by AAAS

of the most carcinogenic materials known (for example, the aflatoxins).

Working with maize, we showed that individual fungal species, which continue to be metabolically active under storage conditions of high relative humidity, replace each other in a process of succession, all the while causing increasing damage within the seeds. We also established that xerotolerant fungi can be transmitted asymptomatically through the growing maize plant, even infecting the next seed generation. Of practical importance in rural Africa was the demonstration that thermotherapy—prestorage immersion of maize grains for short periods in hot water—can substantially reduce fungal loads in maize seeds. Once the seeds are redried after thermotherapy, their potential to be stored for long periods is considerably improved. The reduced mycotoxin levels also make the seeds safer for consumers.

My group's most recent foray into xerotolerant seed fungi centers on the unique gymnosperm of the Namib Desert, *Welwitschia mirabilis*, which is potentially endangered by a seed-associated, ultradesiccation-tolerant fungal species. This plant, which produces only two leaves throughout its long life-span and a root that grows down to the deeply situated water table, provides the only islands of refuge for a variety of small desert animals. Water droplets condensed from nocturnal sea fog run down the plant's long, downward curving leaves, providing the moist environment essential for the animals' survival. Solving the seed fungus problem of *W. mirabilis*, therefore, is of major importance to the survival of the Namib ecosystem.

Peas of a pod. Patricia Berjak (center), surrounded by her collaborators and graduate students, retrieves specimens from cryostorage.

CREDIT: DENISE BERJAK

The Unstorables

In the 1980s my collaborator husband and I, with a succession of our students, began investigations of wet, recalcitrant seeds. Such seeds exhibit unorthodox traits because they cannot withstand dehydration and remain desiccation-sensitive throughout their development and after harvest. The term "recalcitrant," defined as "obstinately disobedient," was first applied by seed scientists to describe the responses of seeds that could not be stored under the conventional low-temperature and low–relative humidity conditions used for orthodox seeds. The category includes seeds of commercially important plants, including those that produce rubber and cocoa, many tropical and subtropical trees, a few temperate species, and a wide spectrum of plants heavily used in Africa for traditional medicine.

Plants in the last-mentioned category include many trees, shrubs, and nonwoody (herbaceous) species of which the bark, leaves, seeds, roots, and bulbs are commodities, collectively worth U.S. $45 million annually. More than 70% of the South African population relies on traditional medicine, and current estimates are that around 4000 tons of plants or plant parts are traded annually in the Durban area alone in traditional medicine (*muthi*) markets.

Many of the plants used for traditional medicine face a double threat—their recalcitrant seeds are short-lived and hard to store, and the plants are overharvested. The pepper-bark tree (*Warburgia salutaris*), for example, has been harvested to extinction in the wild in South Africa.

When we first turned our attention to recalcitrant seeds, little was known about why they could not be dehydrated and why, even if well hydrated, the recalcitrant seeds of most species could be stored only for periods too brief to be useful for long-term conservation of genetic resources. In subtropical Durban, on the eastern seaboard of southern Africa, we were well placed to study seed recalcitrance, having local access to appropriate plant species and the sophisticated laboratory infrastructure necessary to explore the phenomenon.

Using electron microscopy and biochemical analyses, we first showed that highly recalcitrant seeds undergo all the metabolic changes characteristic of the initiation of germination. We showed further that this metabolism continues during the early stages of dehydration, until intracellular damage becomes overwhelming. When Jill Farrant subsequently joined us as a graduate student, we demonstrated that recalcitrant seeds, when stored in hydrated conditions—a humidity high enough to allow the seeds to retain a concentration of water on a par with what it was when the seeds were shed from the tree—become increasingly desiccation-sensitive as the cellular events of germination progress. Without an extraneous water supply, the seeds will begin to deteriorate.

These were definitive discoveries, explaining why visible initiation of germination while seeds were in storage was not merely a nuisance, but was lethal for recalcitrant seeds. A seed that has germinated to the point of requiring additional water will not retain viability unless immediately planted, and is not worth storing. That finding only redoubled our efforts to find new ways of storing these recalcitrant seeds.

Seed Taming

Recalcitrant seeds are not only desiccation-sensitive, but also metabolically active. In contrast, orthodox seeds, owing to their dry state, are metabolically quiescent. Lowering the water content to a level that would preclude germination but facilitate vital metabolism has been suggested as a way to extend the life-span of recalcitrant seeds in hydrated storage. However, Daniel Côme and Françoise Corbineau of the Université Pierre et Marie Curie in Paris, and we, have independently shown that this practice of partial dehydration curtails the seeds' storage life-span.

Current work by graduate students Déon Erdey and Sharon Eggers in our laboratory suggests that slight dehydration stimulates the onset of germinative metabolism, thereby shortening the window of time before additional water is required by the seeds. To optimize storage life-span, just the opposite needs to happen: The onset and progression of germinative metabolism need to be delayed.

Recalcitrant seeds are so-named for a reason. Storage at lowered temperatures might seem an obvious answer, because metabolic rate is slowed in the cold, but many species of tropical and subtropical origin are sensitive to chilling. And even when all the conditions for short- or medium-term hydrated storage

Published by AAAS

have been optimized, most species of recalcitrant seeds face a further limiting factor—fungal infections.

Seed-associated fungi are ubiquitous and pose a prodigious problem: They use seed tissues as their source of nutrition. As a result, the seeds rapidly weaken and die. With graduate student Claudia Calistru and others, we showed that if the seed's fungal load can be reduced or eliminated, then seed storage life-span can be doubled or even quadrupled, depending on the species. Although promising, even this advance is not enough for useful long-term storage of highly recalcitrant seeds.

We also have been pursuing another strategy for halting germinative metabolism in order to increase storage times: deep-freezing. There has long been a consensus that achieving and maintaining the deep-frozen state by cryostorage—usually in liquid nitrogen at –196°C—is the only solution for long-term storage of recalcitrant seeds.

But how can this be achieved practically? Recalcitrant seeds—whether coconuts or the "pips" of a litchi, mango, or avocado pear—are generally large. These seeds also are "wet." Such large, hydrated living structures will not be able to withstand the effects of freezing; the ice crystals wreak lethal havoc on cell structures. The challenge for storing recalcitrant seeds, therefore, depends on both the hydration state and on seed size. The answer lies in reducing both.

We had already begun to make progress in overcoming the obstacles posed by hydration. We previously had observed that the faster the seeds could be dried, the greater the degree of water loss that recalcitrant seeds would tolerate. But large recalcitrant seeds lose water only slowly even under conditions that hasten dehydration. Generally, therefore, the seed dies when its tissues are still too wet to be frozen.

To circumvent that obstacle, H. F. Chin, at the Universiti Pertanian Malaysia, capitalized on a facet of seed anatomy. The bulk of most seeds is made up of tissues containing the nutrient reserves required for germination and seedling growth; only the so-called embryonic axis—the root-shoot continuum, which is very small in most recalcitrant seeds—will ultimately form the new plant. So Chin examined what happens when the embryonic axes, excised from seeds, are quickly dried and frozen. He used commercial rubber tree axes that, although small enough to lose water at a rapid rate, retain viability transiently at the low water contents necessary for successful cryopreservation. Building on Chin's work, my colleagues and I developed a flash-drying technique that permits extremely rapid dehydration of excised axes to water concentrations that allow noninjurious cooling and freezing in liquid nitrogen.

More than 70% of the South African population relies on traditional medicine, and current estimates are that around 4000 tons of plants or plant parts are traded annually in the Durban area alone in traditional medicine (*muthi*) markets.

Flash-drying retains the viability of the seed's embryonic axis at hydration levels close to those of so-called nonfreezable water. Simply put, this is the water that is closely associated with intracellular structures. It does not freeze in any standard sense. In contrast, most of the water within cells occurs as solution water, also called freezable water.

Many of the specific parameters for cryopreservation of excised axes were elucidated and quantified during a collaboration with Christina Walters of the (then) U.S. Department of Agriculture National Seed Storage Laboratory in Fort Collins, Colorado. Together we confirmed that metabolism-linked damage—as opposed to desiccation damage, which occurs when the structure-associated, nonfreezable intracellular water is perturbed—is the basis of seed death during slow dehydration of recalcitrant material. Joined by our then-graduate student, James Wesley-Smith, many of the intricacies of axis survival in relation to drying rate, water concentrations attained, and freezing rate have been—and are still being—resolved.

Axes of temperate seed species have so far proved better able to withstand the procedural "insults" of crystorage than have those of tropical species. These "insults"—axis excision, application of antifungal compounds, dehydration of an essentially desiccation-sensitive structure, plunging into liquid nitrogen, and subsequent thawing and rehydration—would constitute a formidable challenge to any living organism. Nevertheless, encouraging progress has been made, for example, by our colleague Joseph Kioko, whose efforts as a graduate student facilitated drying and successful cryostorage of the seeds of the pepper-bark tree, one of the most sought-after and endangered medicinal plant species in southern Africa.

It is not sufficient that axes merely survive cryostorage: They must ultimately yield growing plants that are phenotypically, genotypically, and physiologically indistinguishable from those grown directly from newly harvested seeds. Among the challenges here are to develop techniques for successful rehydration of axes and for the promotion of shoot production after cryopreservation. We also are developing synthetic seeds—called synseeds—whose individual axes are encapsulated in a gel to reconstitute seedlike structures. This work, performed with postdoctoral fellow Rosa Perán, is in its early stages, and could lead to material that is more easily handled for planting programs.

In time, we hope to offer cryobanking services for recalcitrant-seeded species in Africa. It would be our way of combating the specter of genetic erosion and extinction of the continent's most valuable and sought-after plants.

The author is in the Department of Life and Environmental Sciences, University of KwaZulu-Natal, Durban, South Africa. E-mail: berjak@ukzn.ac.za

10.1126/science.1108429

CREDIT: DEHNVON AHLEFELDT

Published by AAAS

GENOMICS

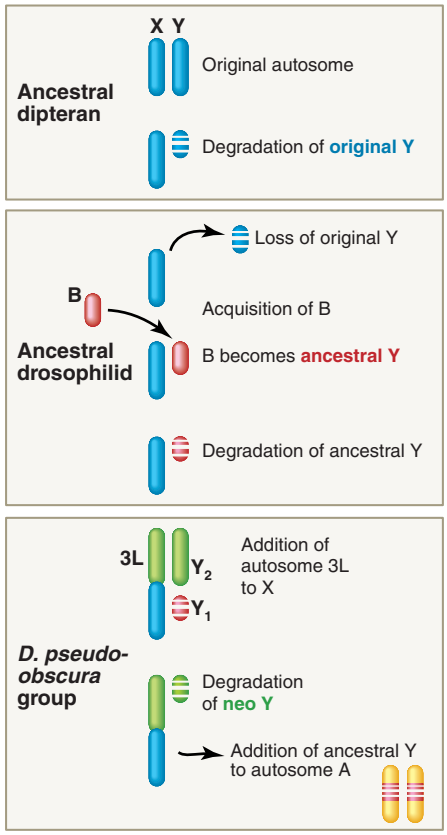
Recycling the Y Chromosome

Jennifer A. Marshall Graves

The Y chromosome must be the strangest element of any genome. In mammals and insects where the mode of chromosomal sex determination is an XX female and XY male, the male-specific Y chromosome is usually small and, unlike ordinary chromosomes, contains numerous repetitive DNA sequences. The Y bears only a few active genes, many of which encode proteins that determine male sex or are involved in reproduction. Comparisons between different species indicate that Y chromosomes degrade rapidly and can even disappear. What happens then?

Fruit flies are model insects that have proved valuable for studying evolutionary changes in sex chromosomes. The original Y chromosome of *Drosophila* species apparently disappeared more than 60 million years ago, and was replaced by a usurper—a blob of repetitive DNA that learned how to pair with the X, and scrounged useful genes from other chromosomes. This "ancestral" Y is shared by species in both major *Drosophila* subgroups. But that is far from the end of the story. Now it appears that in one group of drosophilids, this ancestral Y has itself been retired, and has been recycled by hitching a ride on a non–sex chromosome (autosome). On page 108 of this issue, Carvalho and Clark (*1*) reveal that the Y chromosome of *D. pseudoobscura*—the second of many *Drosophila* species to have its genome sequenced—is unrelated to the ancestral Y. Taking advantage of the outpouring of *D. pseudoobscura* genome data, these investigators tracked down the genes from the ancestral Y of this drosophilid species, and found them leading a blameless existence in ordinary chromosome pairs. Meanwhile, *D. pseudoobscura* sports a new Y, created from an autosome that at some point became fused to the X in this group of flies (see the figure).

Studies of the Y chromosome in many taxa have implied that Y degradation is a one-way street. X and Y chromosomes are thought to have evolved from an ordinary pair of autosomes when one member of the pair acquired a male-determining locus (*2*). Other male-advantage alleles accumulated at this site on the proto-Y, and the male-specific package was preserved by preventing recombination between the X and Y during meiosis. All sorts of dreadful genetic accidents—mutations, deletions, invasion by repetitive elements—occurred in this region of low recombination, progressively degrading the Y into a wasteland of discarded and defunct sequences, and sparing only those genes that acquired a selectable male-specific function. In addition, the Y appropriated some handy genes from autosomes that already had, or later acquired, a male-specific function.



The rise and fall of drosophilid Y chromosomes. The ancestral Y chromosome in most drosophilids, and the neo-Y in the *D. pseudoobscura* subgroup, are completely unrelated. Neither is related to the original Y or to the X, both of which evolved from an autosome pair (blue). The original Y was degraded and ultimately lost. In a drosophilid ancestor 63 million years ago, a B chromosome (red) became paired with the X. This ancestral Y acquired male-specific genes from elsewhere in the genome. In an ancestor of the *D. pseudoobscura* subgroup 18 million years ago, the X fused with an autosome arm (3L, green) to generate an XY_1Y_2 system. The Y_1 chromosome (ancestral Y) fused to an autosome (A, yellow). The Y_2 chromosome underwent partial degradation, retaining only male-specific genes, and now acts as a neo-Y in this group of flies.

This theory accounts well for the peculiarities of the human Y chromosome (*3*). Its ~50-Mb male-specific region retains only 45 active protein-coding genes, many with male-specific functions in sex determination (*SRY*) and spermatogenesis. A few were recruited from other parts of the genome, but most (including those with essential functions in male reproduction) have partners on the X, from which they diverged more than 300 million years ago. These genes are all that remain of the 1400 or so that were on the ancient autosome, now represented by the X. Therefore, we can calculate the rate at which genes have been lost, and predict that the entire human Y will disappear in about 10 million years (*4*). The Y already has been lost in two groups of rodents: the mole voles of Eastern Europe and the spinous country rats of Japan (*5*, *6*). The sex-determination gene *SRY* has evidently been replaced by a new sex-determining trigger elsewhere in the genome. This theory, however, does not account for the Y chromosome of our favorite model insect, *D. melanogaster*, whose Y chromosome is unrelated to the X. Perhaps it was degraded beyond recognition. Or, as with the mole voles, maybe it was completely lost and replaced.

One radical suggestion was that the drosophilid Y evolved, not from the X, but from one of the enigmatic supernumary B chromosomes that sometimes parasitize the genomes of plants and insects and even those of a few mammals. Derived, apparently at random, from all manner of repetitive sequences, B chromosomes live a lonely life, unable to pair with each other or the normal A chromosomes at meiosis. Yet there seems to be some strange affinity between unpartnered nonhomologous chromosomes, prompting the suggestion (*7*) that a B chromosome may have presented itself to the X as a possible partner. Although nonhomologous, this match was the best that either of the unpartnered elements could make. It would confer advantages to both chromosomes at meiosis, where singles are not welcome. Once it began to squire the X during meiosis in male germ cells, this new Y chromosome acquired genes that were moved or copied from autosomes, and these in turn became male-specific. This must have happened at least 60 million years ago, because the ancestral Y is shared by most *Drosophila* species. But not all.

The author is in the Research School of Biological Sciences, Australian National University, Canberra, ACT 2601, Australia. E-mail jenny.graves@anu.edu.au

CREDIT: KATHARINE SUTLIFF/*SCIENCE*

Published by AAAS

The new analysis by Carvalho and Clark (*1*) shows that *D. pseudoobscura* and its close relatives have a completely different Y. This cannot be the original Y that evolved by degradation of the original autosome pair, because the *D. pseudoobscura* subgroup of species diverged only 18 million years ago from the other drosophilid groups and lies nested within species that share the ancestral Y. So this substitution of Y chromosomes must represent a subsequent step and may teach us much about Y chromosome evolution. Was the ancestral Y completely degraded and, in turn, replaced? What happened to useful genes on the ancestral Y? Where did the new Y come from? Carvalho and Clark looked in the genome of *D. pseudoobscura* for homologs of genes on the ancestral Y, and found several of them alive and well and living in clusters on other chromosomes. This suggests that the ancestral Y was not utterly degraded, but was just shifted to the autosomes. Such an event is not unprecedented; at least two genes on the missing Y of the XO spinous country rat have moved to the X (*6*)

Thus, genes that were present in a single copy on the ancestral Y are now present in double-dose copies on an autosome pair. They have retained the testis-specific expression that they evolved during their spell on the ancestral Y. Most strikingly, the genes of the ancestral Y, bloated by millions of years of retroelement insertion, have shrunk back to a normal size in their new location. The ancestral Y has been recycled. There is life after death for the Y chromosome after all.

Where did the new *D. pseudoobscura* Y come from? Carvalho and Clark suggest that it is the relic of yet another autosome. We know that the large X in the *D. pseudoobscura* group was put together by fusion of the X with the arm of an autosome, 3L (see the figure). This leaves not only the original Y (Y_1) unpaired in males, but also one copy of the original autosome (now called Y_2). After recycling of Y_1, the Y_2 chromosome must have taken over, and is now going through the familiar degradation process. Sure enough, Carvalho and Clark found that most of the 15 genes on this neo-Y are homologous to genes on the *D. melangaster* 3L. It will be interesting to see whether these genes have already acquired male-specific functions.

Over the last 100 million years, *Drosophila* species have stuck to the same X, but experimented with at least three quite different Y chromosomes (see the figure). Vertebrate taxa may exhibit a similar phenomenon. For instance, the 5X and 5Y chromosomes of the platypus—which link the mammalian XY and bird ZW sex chromosome systems—exhibit a chain of chromosomes that underwent serial exchanges (translocations) (*8*). Meanwhile, the multiple sex chromosome–autosome translocations found in pygmy mice (*9*) and frogs (*10*) show us how one vertebrate sex chromosome system can evolve into another.

So Y chromosomes can be created and destroyed, re-created and recycled. The Y is dead. Long live the Y.

References

1. A. B. Carvalho, A. G., Clark, *Science* **307**, 108 (2005).
2. B. Charlesworth, D. Charlesworth, *Philos. Trans. R. Soc. London B Biol. Sci.* **355**, 1563 (2000).
3. H. Skaletsky *et al.*, *Nature* **423**, 825 (2003).
4. J. A. M. Graves, *Trends Genet.* **18**, 259 (2002).
5. W. Just *et al.*, *Nature Genet.* **11**, 17 (1995).
6. Y. Arakawa *et al.*, *Cytogenet. Genome Res.* **99**, 303 (2002).
7. J. H. P. Hackstein *et al.*, *BioEssays* **19**, 317 (1996).
8. F. Grützner *et al.*, *Nature* **432**, 913 (2004).
9. F. Veyrunes *et al.*, *Chromosome Res.* **12**, 369 (2004).
10. M. Ogata *et al.*, *Genetics* **164**, 613 (2003).

10.1126/science.1107295

ASTRONOMY

How Is the Solar Corona Heated?

Robert W. Walsh

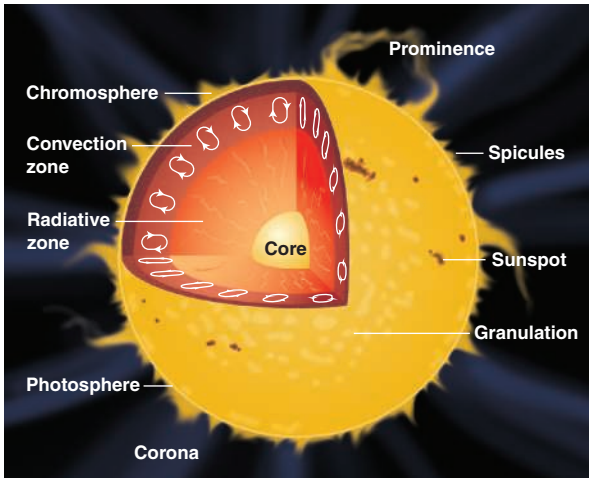
The Sun's outer atmosphere (the corona) contains highly ionized atoms that can only form at a temperature of millions of degrees. Such high temperatures are thermodynamically inexplicable, given that the underlying photosphere and chromosphere (see the first figure) are much cooler (6000 K and at least 4300 K, respectively). Some extra source must deposit energy into the corona to create the steep temperature gradient. This source must also balance energy losses through radiation emitted into space, heat loss from the corona through thermal conduction and mass loss through the solar wind and eruptions such as solar flares. The energy budget needed to power the corona is quite small (about 0.01%) relative to the global solar energy output, but the specific coronal heating mechanism(s) that operate in the corona have been difficult to identify.

Enhanced online at www.sciencemag.org/cgi/content/full/307/5706/51

The author is in the Department of Physics, Astronomy, and Mathematics, University of Central Lancashire, Preston, Lancashire PR1 2HE, UK. E-mail: rwwalsh@uclan.ac.uk

Recently, our understanding of coronal physics has been revolutionized by the space-based Solar and Heliospheric Observatory (SOHO). Residing 1.5 million km sunward from Earth, SOHO's unique vantage point enables continuous monitoring of solar activity. Together with the Transition Region and Coronal Explorer (TRACE) satellite, SOHO can probe the corona at optical, ultraviolet, and extreme ultraviolet wavelengths. The two spacecraft have provided unparalleled observational data, revealing a corona that is dominated by the Sun's magnetic field. Extreme ultraviolet observations show that this field channels the coronal material along the field lines to form a range of loop-like plasma strands (see the second figure).

Because the corona is a hot, magnetized plasma, one should calculate all the forces that act on each charged particle, including the effect of the magnetic field itself. This task is enormous because of the huge number of ions and electrons that make up the plasma. However, the spatial scales considered are often much larger than the dis-

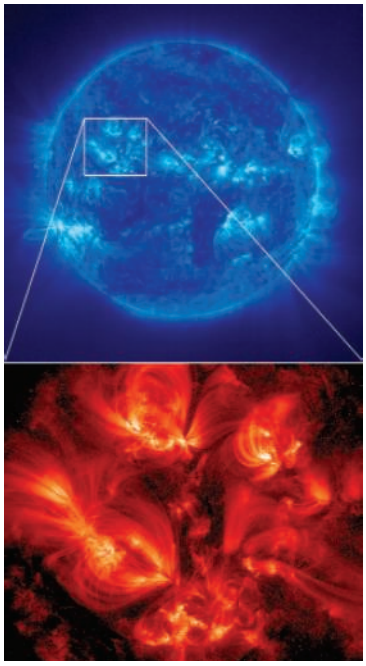


CREDIT: PRESTON HUEY/*SCIENCE*

The structure of the Sun. The solar interior is composed of the core, the radiative zone, and the convection zone; the solar atmosphere consists of the photosphere, the chromosphere, and the corona. Continually streaming out from the Sun into interplanetary space is the solar wind. This sketch is adapted from an educational resource provided by the Space Weather Center (www.spaceweathercenter.org/resources/05/solarscapes/fig1.pdf).

tances between individual charged particles, and the time scales are longer than typical particle collision times. One can therefore treat the corona as a magnetized fluid. Bulk plasma properties (such as temperature, density, pressure, and velocity) can then be investigated using a magnetohydrodynamic approximation. This approach has revealed the evolution of coronal features such as loops and has provided insight into the coronal heating problem.

It is now accepted that the solution to the Sun's counterintuitive temperature gradient lies in the energy that can be transferred to, stored in, and released from the coronal magnetic field. Under normal coronal conditions of high thermal conductivity, magnetohydrodynamics tells us that the magnitude of the magnetic field pressure exceeds that of the corresponding plasma pressure. A "frozen-in-field" condition results, meaning that the magnetic field lines and the plasma advect together when either is subjected to external forcing. This can happen within the magnetic strands, because they are rooted in the turbulent motions of the photosphere and, as a result, subtle magnetic field–plasma interactions can occur. These interactions depend critically on the time it takes for information to travel along the magnetic field (known as the Alfvén time), relative to the time scale over which the field is being displaced (the driving time).



The solar corona at extreme ultraviolet wavelengths. (**Top**) Full-disc image of the Sun's outer atmosphere taken on 20 September 2001 by the extreme-ultraviolet imaging telescope onboard SOHO. The image has a resolution of about 1800 km per pixel side and is at a wavelength of 171 Å, where iron radiates at ~1.3 million K. (**Bottom**) A higher spatial resolution image (about 360 km per pixel side) taken simultaneously at the same wavelength by TRACE. The image displays four strong magnetic field regions with loop-like strands. These loops are believed to follow the magnetic field lines closely, providing a glimpse of the three-dimensional structure of the coronal magnetic field.

There are thus two possible mechanisms for coronal heating: The energy in the corona is dissipated either via propagating magnetohydrodynamic waves (if Alfvén time > driving time) or through small-scale random electric current sheets created within a slowly stressed three-dimensional magnetic field (if Alfvén time < driving time). In both cases, the dissipative effects only become important at length scales of tens of meters or less.

In support of the first mechanism, it is likely that energy-carrying waves are ejected from the photospheric boundary into the corona; the corona is a dynamic environment containing magnetic loops that could support periodic oscillations. The main question is how to dampen and hence extract the energy from the resulting magnetohydrodynamic waves. Much effort has gone into examining cases in which the wave-driving frequency matches some resonance condition along the magnetic field lines, somewhat akin to plucking a stringed instrument.

This could, for example, happen via a process known as phase-mixing. Suppose you have a vertical magnetic field stretching out from a horizontal photospheric boundary. If the Alfvén time varies from one individual field line to the next (for example, due to a density stratification in the horizontal direction), waves that propagate along the field in the vertical direction would quickly become out of phase with one another, creating steep gradients across the wave front that dissipate the wave energy. Similarly, an externally driven, dense plasma strand within the less dense corona may act like a classical oscillator (a process referred to as resonant absorption). Many other wave modes could also be generated (*1*).

The existence of periodic oscillations has been confirmed recently with observational data. In extreme-ultraviolet images taken just tens of seconds apart, both transverse (*2*) and longitudinal (*3*) oscillations in loops have been detected. However, theoretical estimates indicate that waves with much shorter periods (less than 10 s) are required to provide an energy flux that balances coronal losses (*4*). The fact that wave motions with longer periodicity have been found in the corona provides hope that oscillations with shorter periods are also present and that we merely lack the instrumental capabilities to observe them.

Nevertheless, a second possible heating mechanism must also be considered. In the magnetohydrodynamic approximation for the corona, the magnetic diffusion term is usually small (but nonzero). However, this term can become important if the length scales are reduced dramatically. The frozen-in-field condition provides the possibility that oppositely directed magnetic fields may be swept together into a small region, resulting in a large magnetic gradient across a short distance. Field lines within this complicated field structure can then diffuse through the plasma, reconnecting with other field lines to form simpler magnetic topologies. The energy stored in the field is released by this reconnection process in the form of plasma flows, particle acceleration, and heat (*5*).

Therefore, a second possible heating mechanism to consider is the complex entanglement and continual stressing of the magnetic field, which leads to numerous, discrete, localized bursts of energy throughout the corona. The cumulative effect of these "nanoflares" (with energies of about 10^{16} to 10^{17} J per event, compared to 10^{25} to 10^{26} J for a typical solar flare) could provide most of the coronal heating (*6*). Three-dimensional magnetohydrodynamic simulations of braided magnetic fields appear to reproduce such small-scale bursts of energy (*7*), but there is a constraint that nanoflares must overcome.

The frequency distribution (F) of the energy (E) of much larger flares follows a power law: $F(E) = F \times E^{-\alpha}$ with $\alpha = 1.8$. However, dominant nanoflare heating requires $\alpha > 2.0$ (*8*). Statistical calculations of extreme-ultraviolet, small-scale, transient brightenings across the solar disc have been inconclusive, yielding values for α above (*9*) and below (*10*) the critical threshold. This may be because these events are at the edge of our present observational resolution, and because the power-law results appear to be sensitive to how one decides when an explosive brightening has taken place.

It must be remembered that any coronal heating mechanism can only be observed indirectly; what we actually observe is the radiation resulting from the heat that is liberated by a specific process. Therefore, several authors have adopted a forward modeling approach, in which simulated temperatures, densities, and velocities from a magnetohydrodynamic plasma strand model are used to predict an observed quantity (such as the intensity of a given extreme ultraviolet wavelength). This approach has proved successful in mimicking what is observed; for example, such simulations have reproduced to some extent the observed complexity of the plasma's fine structure (*11*).

Building on the success of SOHO and TRACE, the solar physics community has planned several satellite missions over the next decade. These include Solar-B (to study the linkage of the magnetic field though the solar atmosphere); STEREO (in which two spacecraft will drift apart, allowing for stereoscopic reconstruction of the

corona); the Solar Dynamics Observatory (to study solar activity with full-disc, TRACE-quality images captured every 10 s in a range of wavelengths), and Solar Orbiter (which will match the solar rotation and journey out of Earth's ecliptic plane to view the solar poles). The spatial, temporal, and spectral resolution of the spectrometers and atmospheric imagers used by these missions (up to an order of magnitude higher than those of today's instruments) will be vital for finally solving the puzzle of how the solar corona is heated.

References

1. R. W. Walsh, J. Ireland, *Astron. Astrophys. Rev.* **12**, 1 (2003).
2. V. M. Nakariakov *et al.*, *Science* **285**, 862 (1999).
3. I. De Moortel *et al.*, *Astron. Astrophys.* **355**, L23 (2000).
4. L. J. Porter *et al.*, *Astrophys. J.* **435**, 502 (1994).
5. E. R. Priest, T. Forbes, *Magnetic Reconnection: MHD Theory and Applications* (Cambridge Univ. Press, Cambridge, 2000).
6. E. N. Parker, *Astrophys. J.* **330**, 474 (1988).
7. K. Galsgaard, A. Nordlund, *J. Geophys. Res.* **101**, 13445 (1996).
8. H. S. Hudson, *Solar Phys.* **133**, 357 (1991).
9. A. O. Benz, S. Krucker, *Astrophys. J.* **568**, 413 (2002).
10. M. J. Aschwanden *et al.*, *Astrophys. J.* **535**, 1027 (2000).
11. B. V. Gudiksen, A. Nordlund, *Astrophys. J.* **572**, L113 (2002).

10.1126/science.1105460

APPLIED PHYSICS

The Material Is the Machine

Kaushik Bhattacharya and Richard D. James

For more than a century, materials scientists have studied the micrometer-scale needles and platelets that occur in many materials. In martensitic materials—which undergo a reversible, diffusionless solid-to-solid phase transformation in which the underlying crystal lattice spontaneously distorts—the patterns of microstructure resemble a jigsaw puzzle (see the first figure). Recent studies suggest that the characteristic distortions of such martensitic materials can be exploited to create tiny machines.

Imagine that each piece of the puzzle undergoes one of several characteristic distortions. For example, take a crystal with atoms arranged on a cubic lattice and focus on a single cube (a unit cell) with atoms at each corner. In a simple cubic-to-tetragonal transformation, this cube spontaneously elongates along one cube edge. But, by symmetry, it could also elongate along another cube edge. In this case, there are three characteristic distortions and thus three variants of martensite. Martensitic materials have the remarkable property that the puzzle fits together even after undergoing the characteristic distortion, thereby achieving a complex but coordinated motion.

The typical size of the individual domains—the pieces of the jigsaw puzzle—can range from nanometers to millimeters and is determined by a complex interplay between the bulk and interfacial energies (*1*–*4*). This length scale is of interest to researchers trying to make smaller and smaller devices. These researchers have used microelectromechanical systems technology to create intricate patterns on silicon films and multilayers that reproduce, in scaled-down form, the gears, levers, cantilevers, and electromagnetic motors of everyday machines.

K. Bhattacharya is in the Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA. E-mail: bhatta@caltech.edu R. D. James is in the Department of Aerospace Engineering and Mechanics, University of Minnesota, Minneapolis, MN 55455, USA. E-mail: james@umn.edu

This approach has led to devices that can perform unique tasks. An example is digital light projection (*5*), where hundreds of mirrors on a single chip are moved independently to create images on a television or movie screen. But each mirror is made of multiple moving parts that are fabricated on silicon and driven by electrostatic actuators. The complexity thus increases with decreasing size, and there are inherent limitations on how small these devices can reliably be made. Martensitic materials may overcome this limitation. By depositing thin films of a martensitic material and patterning it appropriately, one can control the characteristic distortions to make the domains perform as the components of the machine.

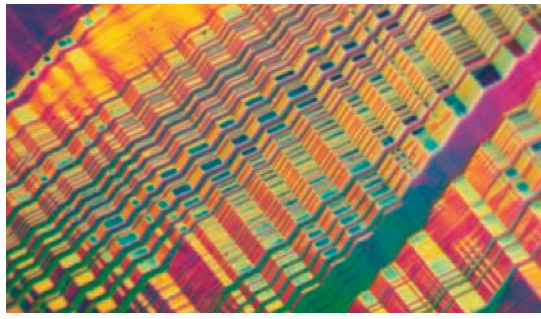
Another potential advantage of martensitic materials was suggested by Krulevitch *et al.* (*6*). Among a broad array of actuator systems, the martensitic material NiTi exhibits the largest known work output per unit volume of the actuator. This performance results from a unique feature of martensitic materials: The linear transformation that defines the distortion of each domain is exactly the same as that of the atomic unit cell of the lattice. The free energy of a domain equals the free energy of its smallest unit cell times the number of unit cells in that domain. A large fraction of this free energy can be transferred to the environment via interaction with a loading device.

A martensitic material thus provides a direct link between the macroscopic environment and its fundamental unit cell. Devices based on this property, such as micropumps, microvalves, and micropositioners, are now reaching commercialization (*7*–*9*). They use polycrystalline films on flexible substrates and exploit the properties averaged over the many domains. Their success motivates the more ambitious quest to use the individual domains as machines.

The microstructure of bulk martensitic materials can be predicted from theory, but until recently this was not possible for thin films of these materials. A powerful new mathematical method, called Γ convergence (*10*), can answer precisely such questions. The resulting theory for thin-film martensites (*11*) has led to a surprising prediction: The interfaces between phases (or, more precisely, between the variants of martensite) in a thin film are completely different from those in bulk material and have a much simpler structure.

This prediction paves the way for the design of a machine: One must pattern the film such that it is released along the predicted interfaces (see the second figure, left panel). In effect, one thus defines the jigsaw puzzle by patterning. This patterning must use the compatible interfaces between phases to select a unique puzzle with a useful distortion.

A proof-of-principle of these concepts has been demonstrated in a single-crystal thinned foil of CuAlNi by Cui and James (see the second figure, right panel) (*12*). The orientation of the foil was chosen so that orthogonal interfaces predicted by the thin-film theory would separate the phases. The foil was then confined outside of a square bounded by these interfaces. This arrangement defined a rather simple jigsaw puzzle with four triangular pieces, which stood up like a tent when cooled and collapsed flat upon heating.



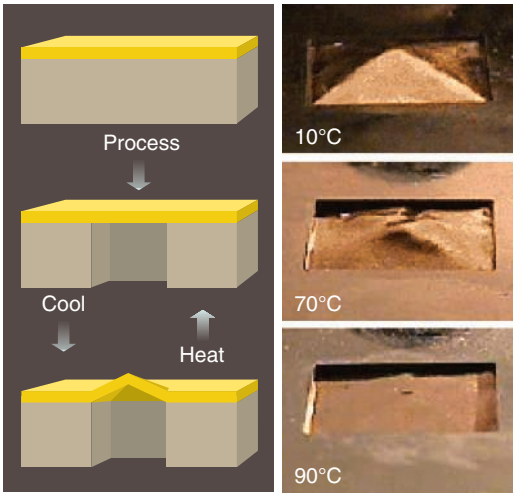
Martensite jigsaw puzzle. Bulk CuAlNi contains six different variants of orthorhombic martensite that form a jigsaw puzzle. The horizontal dimension of this image is 1.4 mm.

In a similar manner, a film released on a strip defined by the predicted interfaces should form a "tunnel." Upon heating, the tunnel would collapse onto the substrate. One can envisage vast networks of such tents and tunnels, with collapsed and un-

CREDIT: C. CHU/NOTRE DAME DE NAMUR UNIVERSITY

collapsed regions driving bits of fluid around. Such networks could be used in microfluidics devices such as labs on chips.

Recent efforts are focusing on enabling this general process on the micrometer and smaller scales. This presents a huge challenge, requiring the synthesis of single-crystal films of martensitic and closely related materials that have complex crystal structures and nonstoichiometric compositions. Recently, researchers armed with a variety of methods have grown such films of perovskite oxides (*13*–*15*) and alloys related to Ni_2MnGa (*16*). (The former are ferroelectrics, whereas the latter alloys are both martensitic and ferromagnetic.) These materials have another interesting feature for material-as-machine: In addition to the distortion, the individual domains are also electrically or magnetically polarized. Thus, instead of causing the shape change by heating or cooling, one can do it by applying a magnetic or electric field, opening up a host of possibilities including remote actuation. Early results are encouraging. For example, Nagarajan *et al.* (*17*) have shown that the domains can be altered by patterning the film and that this method can be exploited to build small-scale actuators.



How materials may act as machines. (**Left**) Schematic cross section of a film, released on a square, showing a tent-like deformation produced by phase transformation. (**Right**) Realization of such a deformation in a CuAlNi foil, released on a 1 cm by 1 cm region, produced by heating and cooling the film.

Given their huge work output per volume and their small scale, materials-as-machine may be best suited for biomedical applications. At these microscopic scales, the forces needed to overcome the enormous constraining effects of both surface tension and viscosity are daunting for microelectromechanical systems. The ideal machine for such applications might in fact be made of proteins.

It turns out that nature already uses such a material-as-machine. Bacteriophage T-4 virus has a tail sheath made of a single-domain protein that undergoes a martensitic transformation (*18*). During invasion of the host, virus-host interactions trigger the transformation in the sheath, driving a hollow protein needle through the cell wall of the bacterium through which its DNA is passed. It would be fascinating to create a human-made analog of the virus's tail sheath. Recent progress in the theory, synthesis, and fabrication of martensitic materials suggests that this idea is ready to explore.

References

1. K. Bhattacharya, *Microstructure of Martensite: Why It Forms and How It Gives Rise to the Shape-Memory Effect* (Oxford Univ. Press, Oxford, 2003).
2. K. Bhattacharya, S. Conti, G. Zanzotto, J. Zimmer, *Nature* **428**, 55 (2004).
3. X. Huang, G. Ackland, K. Rabe, *Nature Mater.* **2**, 307 (2003).
4. G. B. Olson, W. S. Owen, *Martensite* (ASM International, Materials Park, OH, 1992).
5. P. F. Van Kessel, L. J. Hornbeck, R. E. Meier, M. R. Douglas, *Proc. IEEE* **86**, 1687 (1998).
6. P. Krulevitch *et al.*, *J. MEMS* **5**, 271 (1996).
7. A. Ishida, A. Takei, S. Miyazaki, *Thin Solid Films* **228**, 210 (1993).
8. D. L. Polla, L. F. Francis, *Annu. Rev. Mater. Sci.* **28**, 536 (1998).
9. R. H. Wolf, A. H. Heuer, *J. MEMS* **4**, 206 (1995).
10. A. Braides, *Gamma-Convergence for Beginners* (Oxford Univ. Press, Oxford, 2002).
11. K. Bhattacharya, R. D. James, *J. Mech. Phys. Solids* **47**, 531 (1999).
12. J. Cui, R. D. James, in preparation.
13. V. Nagarajan, C. S. Ganpule, R. Ramesh, *Top. Appl. Phys.* **93**, 47 (2004).
14. S. W. Boland, S. C. Pillai, W. D. Yang, S. M. Haile, *J. Mater. Res.* **19**, 1492 (2004).
15. Y. B. Park, J. L. Ruglovsky, H. A. Atwater, *Appl. Phys. Lett.* **85**, 455 (2004).
16. J. W. Dong *et al.*, *J. Appl. Phys.* **91**, 2593 (2004).
17. V. Nagarajan *et al.*, *Nature Mater.* **2**, 43 (2003).
18. G. B. Olson, H. Hartman, *J. de Phys.* **43**, C4-855 (1982).

10.1126/science.1100892

EVOLUTION

Policing Insect Societies

Francis L. W. Ratnieks and Tom Wenseleers

The London Bobby is a reassuring symbol of civic order (*1*). But this symbol is also a reminder that human societies have conflicts. Insect societies, too, experience internal conflicts (*2*), and research increasingly shows that policing is important to resolve these. Within both human and insect societies, conflicts arise because the interests of individuals differ. In insect societies, conflict revolves around reproduction. Reproducing individuals gain by being more closely related to the young males and queens reared in their colony. By reproducing, society members also exploit the colony and this can be costly. First, uncontrolled reproduction upsets the division of labor between queen and workers and results in a less efficient colony. Second, the offspring reared are often genetically less related and so are less valuable to other society members.

To prevent exploitation, social insects have evolved several methods of policing. The best known is "worker policing," whereby workers kill eggs laid by other workers. This phenomenon was first discovered in the honeybee 15 years ago (*3*). Since then, it has been discovered in more than 15 species of bee, wasp, and ant. This past year alone, five more insect species—two species of British wasp (*4*) and three ant species from Florida (*5*), Brazil (*6*), and Finland (*7*)—have been added to the list.

In addition to reducing reproduction by workers, policing also acts to regulate the development of females into distinct queen and worker castes, and to prevent excess females from developing into queens (*8*). When different species are compared, one important overall conclusion emerges: More effective policing results in fewer individuals acting selfishly. There are other striking parallels to human society: Insect policing relies on both detection and prevention, and individuals sometimes attempt to evade policing (see the figure).

In the life of any female bee, wasp, or ant, there are two points at which she may try to reproduce. The first is when, as a larva, she starts developing into either a queen or a worker. In most species, queens are morphologically specialized for egg laying and are often incapable of working. The second is when, as an adult worker, she decides whether to activate her ovaries to lay eggs. In most species, workers cannot mate yet retain ovaries. Therefore, they can lay unfertilized eggs, which develop into males if reared.

Young female larvae of bees, wasps, and ants are usually totipotent, that is, they have the potential to develop into either a queen or a worker. A larva is often better off developing into a queen, yet policing

The authors are in the Laboratory of Apiculture & Social Insects, Department of Animal & Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK, and the Institute of Advanced Study, Berlin 14193, Germany.

CREDIT: (RIGHT PANEL) J. CUI/UNIVERSITY OF MARYLAND

ensures that most are prevented from doing so. Because queens are generally larger than workers and need more food, adult workers can control whether a larva will develop into a queen by controlling her food supply. Consider the honeybee (*8*). A colony reproduces by swarming, dividing into several colonies each headed by the mother queen or a newly reared daughter queen. Prior to swarming, a few queens and thousands of workers are reared. Thus, although a few larvae are reared into queens, with a good chance of heading a colony, thousands of their less fortunate sisters are reared into workers. The fitness advantage of developing into a queen is large. If she ends up heading a colony, the workers will rear her offspring, to which she is related by 0.5. But larvae reared into workers would rear the offspring of a sister queen, to which they are related by only 0.15.

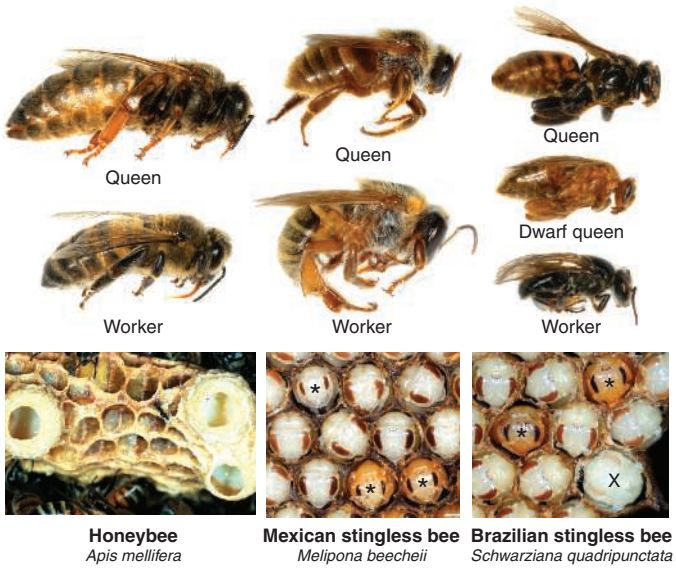
Theory (*8*) shows that about half the honeybee larvae would become queens if they had the power to choose their own caste fate, even though only a few will ever head a colony. Basically, the lottery to reproduce is so attractive that many more enter than could possibly win the prize of heading a new colony. However, the more that enter the lottery, the worse it is for the colony as a whole. This is because queens cannot work. By policing larval development, workers prevent excess queens from being reared and the society operates more efficiently. In essence, what happens is that the workers carefully select and rear a few entrants to the lottery. Entry by all is prevented because there are only a few special large cells in the brood comb in which queens can be raised.

The rearing of queens and workers is so well regulated that only recently has the realization struck that female larvae may try to develop into queens when this is not in their colony's best interest. Stingless bees are the exception that proves the rule. Like honeybees, stingless bees reproduce by swarming and have queens that cannot work. Unlike honeybees, however, they often produce excess queens. Because these queens can serve no useful purpose in the colony, these unfortunate creatures are beheaded or torn apart by the workers soon after they emerge from their cells in the brood comb (*9*). The greatest profligacy occurs in the tropical American genus *Melipona*, with up to 20% of the female larvae developing into queens compared with only about 0.01% in the honeybee (*10*). This excess exists because preventive policing is impossible. Queen and worker *Melipona* have the same body mass (but not shape) and are reared in identical sealed cells, each with its own food supply (see the figure). In contrast, honeybee larvae are reared in open cells and fed progressively, and the cells for queen larvae are bigger than those for worker larvae. As a result, a female *Melipona* larva has enough food to develop into either a queen or a worker and does so while isolated from any intervention by the adult workers. The honeybee, by contrast, has an elegant system of policing, which ensures that only larvae in special queen-size comb cells develop into queens.

Other stingless bees are normally able to prevent excess queen rearing because queens are larger and are reared in special large cells provisioned with extra food. However, larvae reared in smaller cells, which usually would be used to rear workers, sometimes evade their intended fate by developing into dwarf queens. Such dwarf queens are the same size as workers but can still successfully head a colony. Thus, preventing larvae from developing into queens by reducing their food supply is not a fully effective preventive measure in these species. A recent study of one stingless bee species from Brazil, *Schwarziana quadripunctata*, does reveal, however, that dwarf queens are less successful than normal-sized queens, partly because many of the dwarf queens are killed by the workers (*11*). As a result, fewer excess *Schwarziana* queens (about 0.5% of all females) are reared than in the *Melipona* species.

The regulation of queen rearing shows that policing entails prevention, normally making it impossible for larvae to act selfishly by choosing to develop into *beecheii*: a queen rather than a worker (see the figure). The second selfish reproductive strategy, egg laying by workers, demonstrates that policing in insect societies also entails detection: Workers detect and kill worker-laid eggs, thus making it difficult for a selfish worker to profit from her action. A worker bee, wasp, or ant has a strong incentive to lay eggs because she is more related to sons (0.5) than to brothers (queen's sons, 0.25) or nephews (other workers' sons, 0.125 to 0.375). But in many species workers, and sometimes even the queen, kill worker-laid eggs. The effectiveness of this egg policing varies. In the common wasp *Vespula vulgaris* (which nests below ground or in lofts) and the honeybee, almost all (about 98%) of the worker-laid eggs are killed, whereas in *Dolichovespula* wasps (which build football-shaped nests in the open) only 80 to 90% are killed. When policing is more effective, the benefit of laying eggs diminishes so that fewer workers lay eggs. In keeping with this prediction, the proportions of egg-laying workers are much lower in honeybee (0.001 to 0.01%) and common wasp colonies (1%) than in *Dolichovespula* (10%) (*12*).

Egg-laying workers sometimes can evade policing. Eggs laid by anarchistic and parasitic Cape honeybee workers are less closely policed than normal worker-laid eggs (*13*, *14*), probably because they chemically mimic eggs laid by the queen. In ants, workers have another potential means for evading policing. Because ants keep their eggs in piles, it was thought that workers could make their eggs smell queenlike by placing them in the pile. A recent study, however, found that mingling with the queens' eggs offers little protection, perhaps because the queen marks her eggs with



An ounce of prevention... .(**Far left**) In the honeybee *Apis mellifera*, queens are larger than workers and are reared in large, cup-shaped cells at the bottom of the brood comb (extreme left and right). Worker larvae are reared in smaller, hexagonal cells. Adult workers control queen development by ensuring that only queen larvae receive generous food rations, in this way preventing worker larvae from becoming queens. (**Center**) In the Mexican stingless bee *Melipona beecheii*, queens and workers are the same size, leaving the option to develop into a queen open to all female larvae. However, as shown by the piece of uncapped comb (bottom), about 20% selfishly choose to develop as queens (*). Most of these queens are subsequently killed by the workers. This shows that lack of prevention can lead to a large society-level cost. (**Right**) A queen, dwarf queen, and worker of the Brazilian stingless bee *Schwarziana quadripunctata*. The normal queen is larger than a worker but the dwarf queen is not; dwarf queens (*) are reared in the same-size capped cells as workers, whereas normal queens are reared in special cells that are larger in diameter (x) and are located at the comb periphery. By escaping from an intended worker fate, dwarf queens illustrate a strategy of evasion.

chemicals that are not readily transferred (*15*).

Both insects and humans police their societies. Are there any common principles? Can humans learn anything from insect policing? The principal lesson seems to be that policing is a common feature of social life and helps to resolve the conflicts caused by the transition from individuals to societies (*16*). Bees, wasps, and ants each evolved eusociality independently, thus policing must have evolved multiple times. One challenging research question in human policing concerns the effectiveness of different intensities and methods of policing (*17*). This is important because human policing is costly. One 1988 Australian study estimated the annual cost of one additional full-time (24/7) patrol car at A$300,000 (*18*). By showing that individuals are less likely to exploit society when policing is more effective, studies of insects indicate that effective policing can induce individuals to act in ways that are better for society. In insect societies, this means that policing consolidates the basic inequality between queens and workers. Policing in human societies has been used by repressive regimes to sustain inequalities, as demonstrated by the negative connotation of the phrase "Police State." But a human society in which policing is used to promote greater equality and justice may not be an unattractive prospect.

References

1. The Proceedings of the Old Bailey. *Policing in London Before the Bobbies*. www.oldbaileyonline.org/history/crime/policing.html.
2. A. F. G. Bourke, N. R. Franks, *Social Evolution in Ants* (Princeton University, New Jersey, 1995).
3. F. L. W. Ratnieks, P. K. Visscher, *Nature* **342**, 796 (1989).
4. T. Wenseleers *et al.*, *Evolution*, in press.
5. A. Endler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2945 (2004).
6. P. D'Ettorre *et al.*, *Proc. R. Soc. London. Ser.* B **271**, 1427 (2004).
7. H. Helanterä, thesis (University of Helsinki, 2004).
8. T. Wenseleers *et al.*, *J. Evol. Biol.* **16**, 647 (2003).
9. T. Wenseleers *et al.*, *Ethology* **110**, 725 (2004).
10. T. Wenseleers, F. L. W. Ratnieks, *Proc. R. Soc. London Ser.* B (Suppl.) **271**, S310 (2004).
11. T. Wenseleers *et al.*, *Proc. R. Soc. London Ser.* B (Suppl.), in press.
12. T. Wenseleers *et al.*, *J. Evol. Biol.* **17**, 1035 (2004).
13. B. P. Oldroyd, F. L. W. Ratnieks, *Behav. Ecol. Sociobiol.* **47**, 268 (2000).
14. S. J. Martin *et al.*, *Nature* **415**, 163 (2002).
15. P. D'Ettorre *et al.*, in preparation.
16. J. Maynard Smith, E. Szathmáry, *The Major Transitions in Evolution* (Freeman, Oxford, 1995).
17. W. Skogan, K. Frydl, Eds., *Fairness and Effectiveness in Policing: The Evidence* (The National Academies Press, Washington, DC, 2004).
18. P. N. Grabosky, "Efficiency and effectiveness in Australian policing" (report no. 16, Australian Institute of Criminology, Canberra, 1988).

10.1126/science.1106934

IMMUNOLOGY

Decoding Calcium Signaling

Monte M. Winslow and Gerald R. Crabtree

Calcium ions (Ca^{2+}) may be the most widely used second messenger molecules in biology. Indeed, Ca^{2+} is essential for early development of organisms, immune responses, and even for remembering what you are about to read. The diverse tasks of Ca^{2+} raise the question of the molecular origins of specific Ca^{2+} responses. Two recent papers in *Science*, by Badou *et al.* on page 117 of this issue (*1*) and by Launay *et al.* (*2*), identify new players that regulate the movement of Ca^{2+} into and out of cells, providing a more complete understanding of Ca^{2+} responses in the immune system.

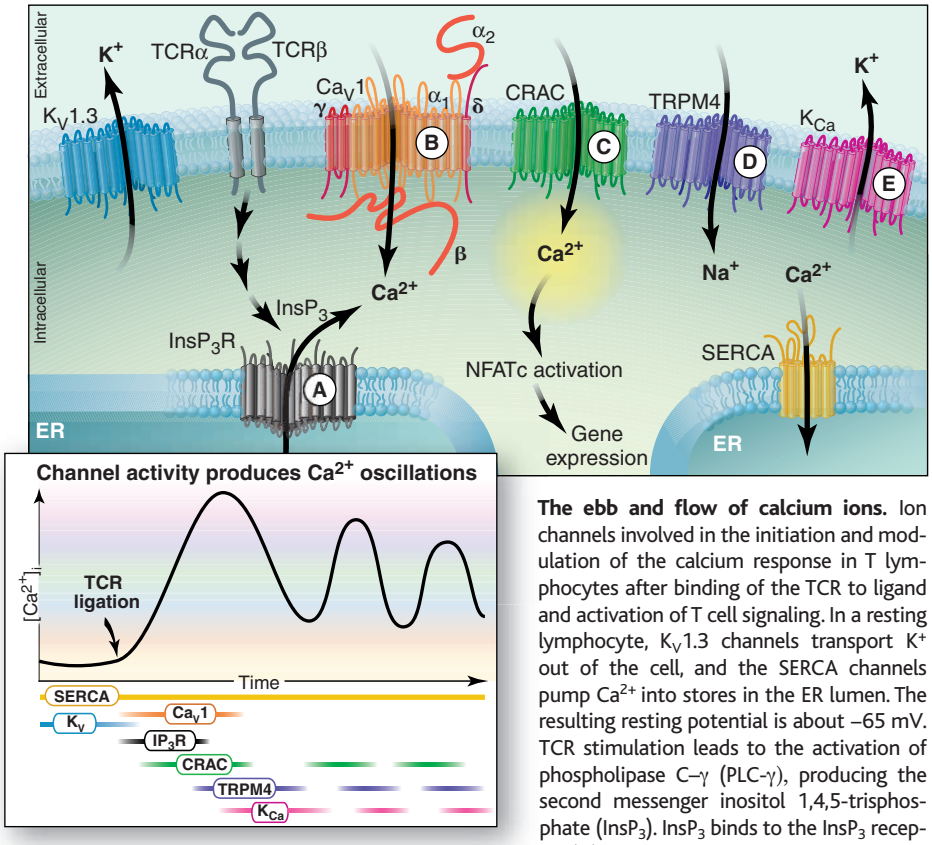
Activation of the T cell receptor (TCR) by ligand initiates a rapid influx of Ca^{2+} into T cells. This leads to the activation of a phosphatase called calcineurin and the subsequent nuclear import and assembly of NFAT transcription complexes, which switch on and off the expression of genes essential for T cell development and function (*3*). Calcineurin is the target of the immunosuppressive drugs cyclosporine A and FK506, the discovery of which ushered in the modern era of organ transplantation. The importance of Ca^{2+}/calcineurin in immune responses and transplantation therapy has generated great interest in how Ca^{2+} signals are regulated. Badou and colleagues examined the *lethargic* mouse (*1*), which carries a mutation in the β_4 subunit of the L-type voltage-gated Ca^{2+} channel (Ca_V1). Previous work showed that L-type Ca^{2+} channels are essential for calcineurin activation and NFAT function in neurons (*4*). Hence it was possible that these channels might also be important for NFAT-dependent transcription in lymphocytes. Indeed, $CD4^+$ T cells from β_4-mutant mice exhibit defects in Ca^{2+} entry in response to TCR signaling, are defective in NFATc1 and NFATc2 dephosphorylation, and exhibit severe defects in cytokine production. But what role do these voltage-gated channels play in the activation of nonexcitable cells such as T lymphocytes?

T cells express many different ion channels in their plasma and endoplasmic reticulum (ER) membranes. These channels coordinately regulate the initial Ca^{2+} spike and sustained Ca^{2+} elevation after antigenic stimulation of T cells. In a resting T cell, K_V channels regulate the flow of potassium ions (K^+) out of the cell, and sarco-endoplasmic reticulum Ca^{2+}-ATPases (SERCA) transport Ca^{2+} into the intracellular stores. The resulting resting potential is about –65 mV. When T cells are stimulated by binding of antigenic peptide associated with major histocompatibility molecules to their TCRs, the lipid second messenger inositol 1,4,5-trisphosphate ($InsP_3$) is generated and triggers the release of Ca^{2+} from intracellular stores through the $InsP_3$ receptor. This increase in cytoplasmic Ca^{2+} concentration ($[Ca^{2+}]_i$) activates the Ca^{2+} release–activated Ca^{2+} (CRAC) channels in the plasma membrane. Once the CRAC channel is activated, the Ca^{2+} influx is driven by the cell's negative resting potential, and the positively charged Ca^{2+} ions flow into the cell. This constitutes the initial rise in $[Ca^{2+}]_i$ observed during the early stages of T cell activation (see the figure).

Badou *et al.* (*1*) provide the first genetic evidence that L-type channels, and specifically β_4-containing L-type calcium channels, may be involved in the initiation of this calcium response in T cells. These channels are composed of α_1, γ, and δ transmembrane subunits as well as an extracellular α_2 and an intracellular β subunit (*5*). L-type Ca^{2+} (Ca_V1) channels have distinct biophysical properties that enable them to participate in signal transduction. The α_1 subunit can associate with any of four distinct cytoplasmic β subunits (β_1 to β_4), all of which may be expressed in T cells (*6*). The β subunit influences the trafficking of the pore-forming α_1 subunit and modulates the characteristics of the channel (*5*).

These investigators suggest that signaling via L-type Ca^{2+} channels influences the initial Ca^{2+} influx based partly on the currents induced after TCR stimulation. The interpretation that these currents are from Ca_V1 channels would be better supported if β_4-mutant T cells failed to produce the TCR-induced currents, which the authors did not test. The presumed coupling of TCR to β_4 raises the possibility that the β_4 or α_1 subunits are regulated by phosphorylation, as has been shown in other cell types, and that this modification serves to facilitate the activation of these channels. Determining the phosphorylation status of the endogenous channel components after antigen receptor signaling may contribute to further understanding of this mechanism. Given that between 10 and a few hundred channels of each type are expressed per T cell, this will not be an easy task. The phenotype of β_4-mutant $CD4^+$ T cells also in-

M. M. Winslow is in the Immunology Program and G. R. Crabtree is in the Departments of Pathology and Developmental Biology at the Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA. E-mail: crabtree@stanford.edu



The ebb and flow of calcium ions. Ion channels involved in the initiation and modulation of the calcium response in T lymphocytes after binding of the TCR to ligand and activation of T cell signaling. In a resting lymphocyte, $K_V1.3$ channels transport K^+ out of the cell, and the SERCA channels pump Ca^{2+} into stores in the ER lumen. The resulting resting potential is about –65 mV. TCR stimulation leads to the activation of phospholipase C–γ (PLC-γ), producing the second messenger inositol 1,4,5-trisphosphate ($InsP_3$). $InsP_3$ binds to the $InsP_3$ receptor (**A**) in the ER membrane, which results in the release of intracellular Ca^{2+}. Ca_V1 channels (**B**) are activated by an undefined mechanism after TCR stimulation and may contribute to the initial increase in $[Ca^{2+}]_i$. Following stimulation of the TCR, Ca^{2+} release–activated Ca^{2+} (CRAC) channels (**C**) open and extracellular Ca^{2+} flows into the cell. When $[Ca^{2+}]_i$ reaches a threshold level, TRPM4 (**D**) is activated and the CRAC channels are inhibited. TRPM4 activation allows Na^+ to enter the cell, which reduces Ca^{2+} flux by depolarizing the membrane. Ca^{2+}-activated K_{Ca} channels (**E**) are also activated, which repolarizes the cells and leads to a subsequent spike in $[Ca^{2+}]_i$. (Inset) Diagram of a typical $[Ca^{2+}]_i$ profile following TCR stimulation (arrow). The channels that are active during each stage are indicated under the profile.

dicates that the other three β subunits cannot substitute for $β_4$. Therefore, specific Ca_V1 channels may have specific functions in T cells. $β_4$-containing Ca_V1 channels are clearly not CRAC channels because Ca^{2+} store-operated entry of Ca^{2+} is normal in $β_4$-mutant $CD4^+$ T cells. This places the Ca_V1 channel in a pathway that is upstream of, or parallel to, that containing the $InsP_3$ receptor. Treatment of T cells with L-type calcium-channel inhibitors indicates that Ca_V1 channels may operate during both the initial and sustained Ca^{2+} influx, hinting at additional roles for the other Ca_V1 channels (*6*, *7*). Patients with mutations in the human ortholog of the Ca_V1 $β_4$ subunit exhibit symptoms of epilepsy and ataxia consistent with the neurological phenotype of the $β_4$-deficient mice (*8*). Analysis of immune cells from these patients also may be informative.

After the initial release of intracellular Ca^{2+} and the activation of the CRAC channel, $[Ca^{2+}]_i$ remains elevated. This sustained elevation of $[Ca^{2+}]_i$ is necessary to regulate gene expression (*9*). As Ca^{2+} levels increase, intracellular Ca^{2+} inhibits the CRAC channel and activates K_{Ca} channels, which transport K^+ out of the cell (see the figure). This combination of negative and positive feedback is important for initiating $[Ca^{2+}]_i$ oscillations with specific characteristics. Launay *et al.* (*2*) provide multiple lines of evidence implicating TRPM4, a Ca^{2+}-activated monovalent cation channel, in Ca^{2+} oscillations. They show that a version of TRPM4 with a truncated amino terminus (ΔN-TRPM4) inhibits sodium ion (Na^+) transport into stimulated lymphocytes. Disruption of TRPM4 activity resulted in an increase in $[Ca^{2+}]_i$ after TCR stimulation without the characteristic Ca^{2+} oscillation, and also increased cytokine production. This ion channel may decrease the electrochemical driving force for Ca^{2+} entry, thereby allowing oscillations in $[Ca^{2+}]_i$. Complementary data from cells in which TRPM4 expression has been decreased by RNA interference (RNAi) strengthen the conclusion that the dominant-negative form of TRPM4 (ΔN-TRPM4) is specific for TRPM4 channels among the large TRP channel family (*10*). However, studies of mice lacking TRPM4 will be required to determine the importance of TRPM4 in vivo. It is possible that TRPM4-deficient lymphocytes might either produce unsurpassed immune responses or deadly autoimmunity.

The results of Badou *et al.* (*1*) and Launay *et al.* (*2*) raise the exciting possibility of developing therapeutics that target these newly described channels. For example, inhibitors of K_v channels attenuate the initial Ca^{2+} influx and early gene induction after T cell stimulation (*11*). Specific and high-affinity inhibitors of the Ca_V1 channels, or of the $β_4$ subunit only, may inhibit lymphocyte activation in vivo. Pharmacological activation of TRPM4 channels would decrease the driving force for Ca^{2+} entry and lead to immunosuppression, whereas TRPM4 inhibitors might lead to more rapid pathogen clearance or more robust responses to immunization. Small changes in channel expression can have profound effects on $[Ca^{2+}]_i$. For example, RNAi of TRPM4 mRNA, which reduces TRPM4 protein levels by half, has a dramatic effect on Ca^{2+} flux after TCR stimulation. Interestingly, the expression of many of these ion channels dramatically changes after lymphocyte activation, indicating that effector and memory T cells are likely to express distinct combinations and different amounts of these ion channels (*1*, *5*). The natural differences in ion-channel expression among different lymphocyte populations may also permit specific cell types to be targeted owing to their unique channel profiles. Whereas this research area will likely become more complex before it becomes clear, the work by Badou *et al.* (*1*) and Launay *et al.* (*2*) provides insight into the most poorly understood step in T lymphocyte activation, and sets the stage for developing practical ways to modulate lymphocyte activity.

References and Notes

1. A. Badou *et al.*, *Science* **307**, 117 (2005).
2. P. Launay *et al.*, *Science* **306**, 1374 (2004).
3. G. R. Crabtree, E. N. Olson, *Cell* **109**, S67 (2003).
4. I. A. Graef *et al.*, *Nature* **401**, 703 (1999).
5. W. A. Catteral, *Annu. Rev. Cell Biol.* **16**, 521 (2000).
6. L. Stokes, J. Gordon, G. Grafton, *J. Biol. Chem.* **279**, 19566 (2004).
7. M. Kotturi, D. A. Carlow, J. C. Lee, H. J. Ziltener, W. A. Jefferies, *J. Biol. Chem.* **278**, 46949 (2003).
8. A. Escayg *et al.*, *Am. J. Hum. Genet.* **66**, 1531 (2000).
9. S. Feske, J. Giltnane, R. Dolmetsch, L. M. Staudt, A. Rao, *Nature Immunol.* **2**, 316 (2001).
10. C. Montell, L. Birnbaumer, V. Flockerzi, *Cell* **108**, 595 (2002).
11. W. A. Schmalhofer *et al.*, *Biochemistry* **41**, 7781 (2002).

12. We thank C. Barrett and R. Lewis for helpful comments and apologize to those who were not cited owing to space limitations. M.M.W. is supported by a Stanford Graduate Fellowship and a Howard Hughes Predoctoral Fellowship. G.R.C. is funded by grants from the NIH and is an Investigator for the Howard Hughes Medical Institute.

10.1126/science.1108163

Normalization of Tumor Vasculature: An Emerging Concept in Antiangiogenic Therapy

Rakesh K. Jain

Solid tumors require blood vessels for growth, and many new cancer therapies are directed against the tumor vasculature. The widely held view is that these antiangiogenic therapies should destroy the tumor vasculature, thereby depriving the tumor of oxygen and nutrients. Here, I review emerging evidence supporting an alternative hypothesis—that certain antiangiogenic agents can also transiently "normalize" the abnormal structure and function of tumor vasculature to make it more efficient for oxygen and drug delivery. Drugs that induce vascular normalization can alleviate hypoxia and increase the efficacy of conventional therapies if both are carefully scheduled. A better understanding of the molecular and cellular underpinnings of vascular normalization may ultimately lead to more effective therapies not only for cancer but also for diseases with abnormal vasculature, as well as regenerative medicine, in which the goal is to create and maintain a functionally normal vasculature.

Solid tumors account for more than 85% of cancer mortality. Because cancer cells in these tumors require access to blood vessels for growth and metastasis, inhibiting vessel formation offers hope for reducing the mortality and morbidity from these tumors (*1*). When administered as single agents, antiangiogenic drugs have produced modest objective responses in clinical trials (*2*, *3*), but overall they have not yielded long-term survival benefits (*4*). In contrast, when given in combination with chemotherapy, bevacizumab, an antibody targeted against the potent angiogenic molecule vascular endothelial growth factor (VEGF), produced an unprecedented increase in survival (5 months) in colorectal cancer patients (*5*).

These clinical data support the earlier predictions of Teicher (*6*), who postulated that combined administration of antiangiogenic and cytotoxic (chemo- and radiation) therapies would yield maximal benefit because such combinations would destroy two separate compartments of tumors—cancer cells and endothelial cells. Cytotoxic agents would kill cancer cells directly, and antiangiogenic agents would kill cancer cells indirectly by depriving them of nutrients. Emerging data suggest that chemotherapy and radiation therapy may also have antiangiogenic effects, directly damaging or killing tumor endothelial cells, bone marrow–derived cells (e.g., endothelial progenitor cells), and/or circulating endothelial cells, and thus enhancing the indirect killing of cancer cells (*7*). Furthermore, cancer cells may express receptors for angiogenic growth factors (e.g., VEGFR1 or VEGFR2), and thus antiangiogenic drugs (e.g., antibody to VEGF) could lead to the direct killing of cancer cells by interfering with survival pathways and/or enhancing sensitivity to other treatments (*7*). All of these mechanisms imply that an antiangiogenic agent would always augment the response to radiation or chemotherapy.

This is paradoxical, however. One would expect that destroying the vasculature would severely compromise the delivery of oxygen and therapeutics to the solid tumor, producing hypoxia that would render many chemotherapeutics, as well as radiation, less effective. Indeed, some studies show that antiangiogenic therapy can compromise the delivery of drugs to tumors (*8*), as well as antagonize the outcome of radiation therapy (*9*, *10*). How can we reconcile these diametrically opposed outcomes of combination therapy with the positive preclinical and clinical data? The answer to this question is critical for optimizing the efficacy of combined antiangiogenic and cytotoxic therapy.

To resolve this paradox, I have hypothesized that the judicious application of antiangiogenic agents can "normalize" the abnormal tumor vasculature, resulting in more efficient delivery of drugs and oxygen to the targeted cancer cells (Fig. 1A) (*11*). The increased penetration of drugs throughout the tumor would enhance the outcome of chemotherapy, and the ensuing increased level of oxygen would enhance the efficacy of radiation therapy and many chemotherapeutic agents. Here, I review our current understanding of vascular normalization, discuss recent preclinical and clinical data supporting this counterintuitive hypothesis, and point out potential research avenues for further exploration.

E. L. Steele Lab for Tumor Biology, Department of Radiation Oncology, Massachusetts General Hospital, and Harvard Medical School, Cox-7, 100 Blossom Street, Boston, MA 02114, USA. E-mail: jain@steele.mgh.harvard.edu

Why Normalize the Tumor Vasculature?

To obtain nutrients for their growth and to metastasize to distant organs, cancer cells co-opt host vessels, sprout new vessels from existing ones (angiogenesis), and/or recruit endothelial cells from the bone marrow (postnatal vasculogenesis) (*12*). The resulting vasculature is structurally and functionally abnormal (Table 1) (*13*). Blood vessels are leaky, tortuous, dilated, and saccular and have a haphazard pattern of interconnection (Fig. 1). The endothelial cells lining these vessels have aberrant morphology, pericytes (cells that provide support for the endothelial cells) are loosely attached or absent, and the basement membrane is often abnormal—unusually thick at times, entirely absent at others.

These structural abnormalities contribute to spatial and temporal heterogeneity in tumor blood flow. In addition, solid pressure generated by proliferating cancer cells compresses intratumor blood and lymphatic vessels, which further impairs not only the blood flow but also the lymphatic flow (*14*). Collectively these vascular abnormalities lead to an abnormal tumor microenvironment characterized by interstitial hypertension (elevated hydrostatic pressure outside the blood vessels), hypoxia, and acidosis.

Impaired blood supply and interstitial hypertension interfere with the delivery of therapeutics to solid tumors. Hypoxia renders tumor cells resistant to both radiation and several cytotoxic drugs. Independent of these effects, hypoxia also induces genetic instability and selects for more malignant cells with increased metastatic potential (*15*). Hypoxia and low pH also compromise the cytotoxic functions of immune cells that infiltrate a tumor. Unfortunately, cancer cells are able to survive in this abnormal microenvironment. In essence, the abnormal vasculature of tumors and the resulting abnormal microenvironment together pose a formidable barrier to the delivery and efficacy of cancer therapy. This suggests that if we knew how to correct the structure and function of tumor vessels, we would have a chance to normalize the tumor microenvironment and ultimately

to improve cancer treatment. The fortified tumor vasculature may also inhibit the shedding of cancer cells into the circulation—a prerequisite for metastasis.

In the past, higher doses of drugs and hyperbaric oxygenation have been used to increase the tumor concentrations of drugs and oxygen, respectively. These strategies have not shown much success in the clinic, however. One reason for this failure is that tumor vessels have large holes in their walls (*16*). As stated earlier, this leakiness leads to interstitial hypertension as well as spatially and temporally nonuniform blood flow. If the delivery system is flawed, it does not matter how much material is pumped into it. The drugs and oxygen will become concentrated in regions that already have enough and will still not reach the inaccessible regions (*17*). However, if we fix the delivery system, more cells are likely to encounter an effective concentration of drugs and oxygen. This is the rationale for developing therapies that normalize the tumor vasculature. These therapies do not merely increase the total uptake of drugs and oxygen but also distribute these molecules to a larger fraction of the tumor cells by fixing the delivery system.

How Should We Normalize the Tumor Vasculature?

In normal tissues, the collective action of angiogenic stimulators (e.g., VEGF) is counterbalanced by the collective action of angiogenic inhibitors such as thrombospondin-1 (Fig. 1D). This balance tips in favor of the stimulators in both pathological and physiological angiogenesis (*18*). However, in pathological angiogenesis, the imbalance persists. Therefore, restoring the balance may render the tumor vasculature close to normal. On the other hand, tipping this balance in favor of inhibitors may lead to vascular regression and, ultimately, to tumor regression.

If we had antiangiogenic agents that completely destroyed tumor vessels without harming normal vessels, we would not need to add cytotoxic therapy. Unfortunately, such agents are not currently available. It is conceivable that increased doses of currently available antiangiogenic agents could produce complete tumor regression, but such doses are likely to adversely affect the vasculature of normal tissues, including the cardiovascular, endocrine, and nervous systems (*12*). Indeed, antiangiogenic therapy with agents such as bevacizumab is associated with an increased risk of arterial thromboembolic events (*19*), and such adverse effects could be more pronounced with increased doses. Furthermore, excessive vascular regression may be counterproductive because it compromises the delivery of drugs and oxygen (Fig. 1). Indeed, suboptimal doses or scheduling of antiangiogenic agents might lower tumor oxygenation and drug delivery and, thus, antagonize rather than augment the response to radiotherapy or chemotherapy (*8*–*10*). This need for a delicate balance between normalization and excessive vascular regression emphasizes the requirement for careful selection of the dose and administration schedule for antiangiogenic agents.

Can Blocking VEGF Signaling Normalize Tumor Vessels?

Of all the known angiogenic molecules, VEGF (also referred to as VEGF-A) appears the most critical (*12*, *20*, *21*). VEGF promotes the survival and proliferation of endothelial cells, increases the display of adhesion molecules on these cells, and increases vascular permeability. During mouse embryonic development, the exquisite regulation of VEGF expression sets in motion a chain of events that leads to the development of a mature vasculature from primordial cells (*18*). Deletion of a single allele of VEGF results in embryonic lethality. So, too, does overexpression of VEGF. In adults, ectopic overexpression of VEGF results in a highly abnormal vasculature (*22*). Collectively, these results indicate that the normal vasculature requires precise spatial and temporal control of VEGF levels.

VEGF is overexpressed in the majority of solid tumors. Thus, if one were to judiciously down-regulate VEGF signaling in tumors, then the vasculature might revert back to a more ''normal'' state. Indeed, blockade of VEGF signaling passively prunes the immature and leaky vessels of transplanted tumors in mice and actively remodels the remaining vasculature so that it more closely resembles the normal vasculature (Fig. 1). This ''normalized'' vasculature is characterized by less leaky, less dilated, and less tortuous vessels with a more normal basement membrane and greater coverage by pericytes (Fig. 1C). These morphological changes are accompanied by functional changes—decreased interstitial fluid pressure, increased tumor oxygenation, and improved penetration of drugs in these tumors (Table 1) (*6*, *16*, *23*–*30*).

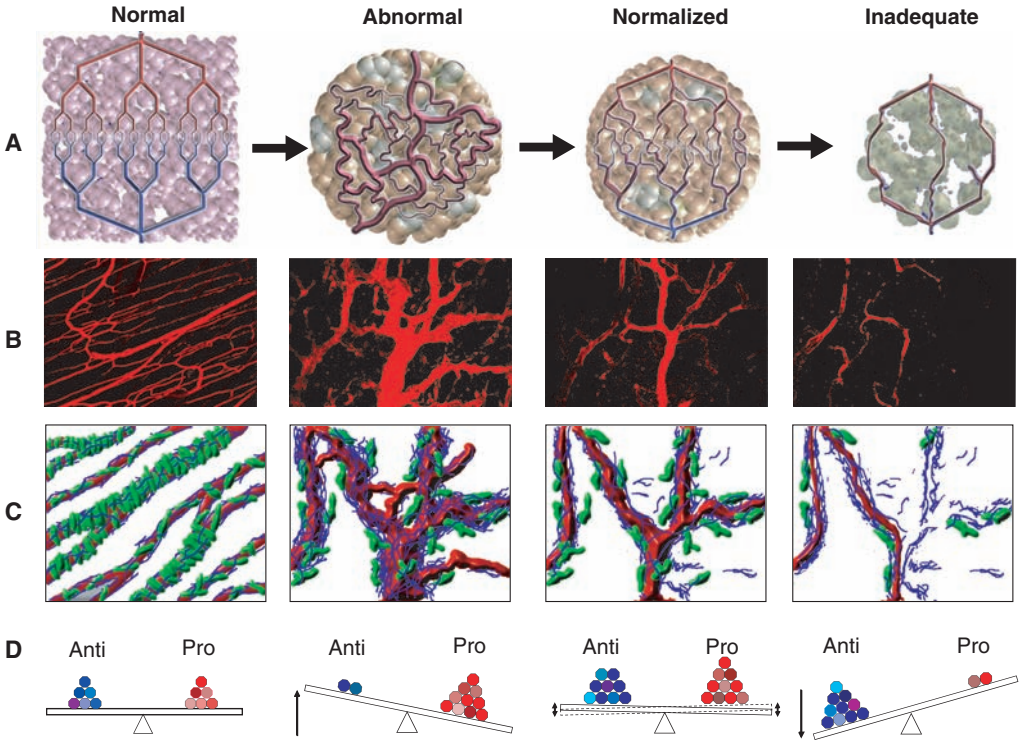


Fig. 1. Proposed role of vessel normalization in the response of tumors to antiangiogenic therapy. (**A**) Tumor vasculature is structurally and functionally abnormal. It is proposed that antiangiogenic therapies initially improve both the structure and the function of tumor vessels. However, sustained or aggressive antiangiogenic regimens may eventually prune away these vessels, resulting in a vasculature that is both resistant to further treatment and inadequate for the delivery of drugs or oxygen [reproduced, with permission, from (*11*)]. (**B**) Dynamics of vascular normalization induced by VEGFR2 blockade. On the left is a two-photon image showing normal blood vessels in skeletal muscle; subsequent images show human colon carcinoma vasculature in mice at day 0, day 3, and day 5 after administration of VEGR2-specific antibody [reproduced, with permission, from (*24*)]. (**C**) Diagram depicting the concomitant changes in pericyte (red) and basement membrane (blue) coverage during vascular normalization (*24*, *29*). (**D**) These phenotypic changes in the vasculature may reflect changes in the balance of pro- and antiangiogenic factors in the tissue.

Table 1. Morphological and functional characteristics of the vasculature in normal tissue, an untreated tumor, a tumor during early stages of treatment with an antiangiogenic drug (normalized), and a tumor treated with high doses of an antiangiogenic drug over a long period (regressing). MVP, microvascular pressure; IFP, interstitial fluid pressure.

Properties	Vessel type				Reference
	Normal	Tumor (untreated)	Tumor (normalized)	Tumor (regressing)	
Global organization	Normal	Abnormal	Normalized	Fragmented	(*30*)
Pericyte	Normal	Absent or detached	Closer to normal	Missing	(*24*, *27*, *29*, *31*)
Basement membrane	Normal	Absent or too thick	Closer to normal, some ghost	Ghost	(*24*, *27*, *29*)
Vessel diameter	Normal distribution	Dilated	Closer to normal	Closer to or less than normal	(*24*, *25*, *29*, *37*)
Vascular density	Normal, homogeneous distribution	Abnormal, heterogeneous distribution	Closer to normal	Extremely low	(*24*–*29*, *31*)
Permeability to large molecules	Normal	High	Intermediate	Variable	(*24*–*26*, *29*, *31*, *37*)
MVP and IFP	MVP > IFP	MVP ~ IFP	MVP > IFP	Low IFP	(*23*, *24*, *31*)
Plasma (P) and interstitial (I) oncotic pressure*	P > I	P ~ I	P > I	Not known	(*24*)
pO_2	Normal	Hypoxia	Reduced hypoxia	Hypoxia	(*6*, *23*, *29*)
Drug penetration	Uniform	Heterogeneous	More homogeneous	Inadequate	(*6*, *24*, *28*, *31*)

*Osmotic pressure exerted by plasma proteins.

What About Human Tumors?

Thousands of patients worldwide have received anti-VEGF therapy. The effect of VEGF blockade on human tumors was recently studied in rectal carcinoma patients receiving an antibody to VEGF, bevacizumab, together with radiation and chemotherapy (*31*). The results in patients mirrored those seen in transplanted tumors in mice: Two weeks after a single injection of bevacizumab alone, the global (mean) blood flow of tumors, as measured by contrast-enhanced computed tomography (CT), decreased by 30 to 50% in six consecutive patients. Tumor microvascular density, vascular volume, and interstitial fluid pressure were also found to be reduced. Surprisingly, however, there was no concurrent decrease in the uptake of radioactive tracers in tumors, which suggests that vessels in the residual "normalized" tumor vasculature were more efficient in delivering these agents to tumor parenchyma than they were prior to bevacizumab treatment.

Similar reductions in blood flow, as measured by magnetic resonance imaging (MRI), had been noted previously in patients treated daily with small-molecule inhibitors of VEGFR tyrosine kinase activity (PTK787 and SU6668) (*32*, *33*). Interestingly, however, positron emission tomography (PET) analysis of patients treated with endostatin, an endogenous inhibitor of angiogenesis, revealed a biphasic response—that is, an increase in tumor blood flow at lower doses and a decrease at intermediate doses (*34*).

As we interpret these human data obtained from MRI, PET, and CT, two key limitations of these imaging methods must be kept in mind. First, because most of these methods yield a parameter that depends on both blood flow and permeability, the blood flow and permeability cannot be calculated unambiguously. Second, tumor blood flow is highly heterogeneous. It is not the total blood flow, but the distribution of blood flow, that determines the distribution of a drug or oxygen in tumors. Therefore, the global (total) blood flow, as estimated by the currently available resolution of MRI, CT, or PET, does not inform us about the degree of spatial heterogeneity in vascular normalization or drug distribution (*17*). Thus, improved imaging techniques, which can measure the spatial and temporal changes in blood flow and other physiological parameters with higher resolution, are needed to definitively establish the effects of antiangiogenic treatment on vascular function in human tumors growing at different sites.

Is There an Optimal Time or Drug Dose for Normalization?

Optimal scheduling of antiangiogenic therapy with chemotherapy and/or radiation therapy requires knowledge of the time window during which the vessels initially become normalized, as well as an understanding of how long they remain in that state. Recent studies, in which human tumors growing in mice were treated with an antibody to VEGF receptor-2, have identified such a "normalization window," that is, a period during which the addition of radiation therapy yields the best therapeutic outcome (Fig. 2) (*29*). This window was short-lived (about 6 days) and was characterized by an increase in tumor oxygenation, which enhances radiation therapy by increasing the concentration of reactive oxygen species created by the radiation. During the normalization window, but not before or after it, VEGFR2 blockade was found to increase pericyte coverage of vessels in a human brain tumor grown in mice. Vessel normalization was accompanied by up-regulation of Angiopoietin 1 and activation of matrix metalloproteinases (MMPs). The prevailing hypothesis is that VEGF blockade passively prunes nascent vessels that are not covered with pericytes. In contrast, this study found that pericyte coverage increased prior to vascular pruning (*29*). Improved understanding of the molecular mechanisms of vessel normalization may suggest new strategies for extending the normalization window to provide ample time for cytotoxic therapy.

The dose of antiangiogenic agents also determines the efficacy of combination therapy. Although it is tempting to increase the dose of antiangiogenic agents or to use a more potent angiogenic blocker, as one would for chemotherapeutic agents, doing so might lead to normal-tissue toxicity and compromise the tumor vessels to the point that drug delivery is impaired. Indeed, renal cell carcinoma patients on a high dose of bevacizumab (10 mg per kg of body weight every 2 weeks) were more likely to develop hypertension and proteinuria than those on a low dose, although the sample size was too small for comparison of the rates of serious adverse events (*2*). Even the low dose of bevacizumab (5 mg/kg) given in combination with chemotherapy has contributed to an increased risk of cardiovascular problems, including death, in some cancer patients (*19*). Although no dose comparison has yet been made in large clinical trials, it is conceivable that such serious adverse events may increase with higher doses. In studies of mice, more potent blockers of VEGF signaling have induced regression of normal tracheal and thyroid vessels (*35*).

Do We Need an Antiangiogenic Cocktail for Normalization?

The constellation of angiogenic molecules expressed in a tumor increases with malignant progression, rendering certain tumors less dependent on VEGF. For example, early stages of breast tumors may require only VEGF for angiogenesis, whereas at later stages, angiogenesis

in these tumors may be driven by additional factors, including fibroblast growth factor 1 (FGF-1), FGF-2, transforming growth factor–β (TGF-β), platelet-derived endothelial cell growth factor (PD-ECGF), and placental growth factor (PlGF) (*36*). Thus, a late-stage breast tumor may escape anti-VEGF treatment by exploiting alternative angiogenic factors to generate its neovasculature. This may help explain why the addition of bevacizumab to chemotherapy did not prolong the survival of breast cancer patients in a recent phase III trial (*4*). Optimal cancer treatment may require the targeting of multiple angiogenic pathways, and the challenge for the oncologist will be to formulate cocktails of antiangiogenic agents specifically tailored to the angiogenic profile of individual tumors.

Small-molecule inhibitors that target multiple kinases involved in tumor angiogenesis (e.g., VEGFR2, PDFGRβ) are currently in clinical trials (*7*). Whether these kinase inhibitors cover the necessary spectrum of angiogenic pathways to normalize the tumor vasculature is not known. If they do not, it will be important to conduct clinical trials that combine antiangiogenic agents produced by competing pharmaceutical companies before each agent is approved by the Food and Drug Administration (FDA).

An alternative approach is to develop agents that mimic such antiangiogenic cocktails by targeting upstream pathways that regulate the production of angiogenic molecules. Such drugs already exist and have been approved by the FDA. For example, one recent study has shown that an antibody to HER2, trastuzumab (also known as Herceptin), acts on multiple angiogenic pathways in HER2-overexpressing human breast cancer xenografts (*37*). Herceptin lowers the expression of several proangiogenic molecules while increasing expression of the antiangiogenic molecule thrombospondin-1. Interestingly, although Herceptin lowered the expression of VEGF in tumor cells, the host cells within the tumor stroma produced compensatory VEGF; thus, additional anti-VEGF treatment could improve the efficacy of Herceptin. On the basis of these preclinical results, a clinical trial has been initiated in which Herceptin is combined with bevacizumab for treatment of HER2-positive breast cancer (*38*).

A further finding from this study was that Herceptin "normalized" the vasculature of human breast cancer xenografts (*37*). Whereas vessels in the control antibody-treated tumors were dilated and leaky, those in the Herceptin-treated tumors had diameters and vascular permeability closer to those of normal vessels. Thus, Herceptin and other drugs that target upstream mutant receptors [e.g., cetuximab, an antibody to HER1 (*39*), or kinase inhibitors such as gefitinib or imatinib (*40*)] act as mimics of antiangiogenic cocktails, that is, these drugs improve their own delivery as well as that of other therapeutics given in combination. This improvement in delivery and alleviation of hypoxia presumably contributes to their efficacy.

These findings point to an urgent need to investigate whether other molecularly targeted agents also mimic antiangiogenic cocktails. This concept may not only change our view of how these therapeutics actually work in vivo but will also provide us with the knowledge of how to use them for improving the delivery of other therapeutic agents given in combination. Furthermore, once we discover the set of angiogenic pathways affected by each drug, we can develop algorithms to combine them for maximal efficacy in a manner tailored to the angiogenic profile of each tumor. Further support for the use of these agents as antiangiogenic cocktails comes from emerging evidence that most oncogene and tumor-suppressor gene pathways are implicated in angiogenesis, either directly or indirectly (*41*–*43*).

Is Tumor Growth Accelerated During Vascular Normalization?

One would expect that the improved delivery of oxygen and nutrients during vascular normalization would enhance tumor growth. However, both preclinical and clinical studies to date show that, despite normalization, tumor growth is not accelerated during antiangiogenic monotherapy. There are several possible explanations for this apparent paradox.

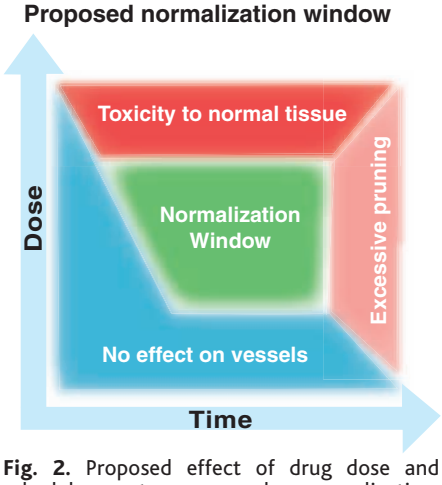


Fig. 2. Proposed effect of drug dose and schedule on tumor vascular normalization. The efficacy of cancer therapies that combine antiangiogenic and cytotoxic drugs depends on the dose and delivery schedule of each drug. The vascular normalization model posits that a well-designed strategy should passively prune away immature, dysfunctional vessels and actively fortify those remaining, while incurring minimal damage to normal tissue vasculature. During this "normalization" window (green), cancer cells may be more vulnerable to traditional cytotoxic therapies and to novel targeted therapies. The degree of normalization will be spatially and temporally dependent in a tumor. Vascular normalization will occur only in regions of the tumor where the imbalance of pro- and antiangiogenic molecules has been corrected.

1) It is important to remember that vascular normalization occurs in the context of antiangiogenic treatment and that the main effect of this treatment is a reduction in the number of blood vessels (vessel density), which should lead to tumor regression. Moreover, tumors are highly heterogeneous; not all regions are equally vascularized, some tumor vessels are more mature than others, and the balance of pro- and antiangiogenic molecules differs from region to region and from one moment to the next. Hence, it may be that the effects of vessel normalization in some regions of the tumor are swamped by simultaneous vessel regression in other regions. In addition, the inability of tumors to grow new vessels during antiangiogenic therapy limits the ability of this transient increase in vascular efficiency to expand the tumor mass. If it were easy to achieve complete tumor regression with antiangiogenic monotherapy, vascular normalization would be of marginal importance, because it is expected to affect only a subset of cells and to do so only temporarily. Unfortunately, some tumor cells are able to survive antiangiogenic monotherapy, and these cells must be targeted with combined therapy.

2) The transient normalization of tumor vessels produces a temporary increase in oxygen and nutrient delivery to the cancer cells that surround these "normalized" vessels. This might be expected to enhance the proliferation of these cells and hence to accelerate tumor growth. However, this intuitive notion is not supported by published data. For example, Gullino found no correlation between tumor growth rate in vivo and blood flow rate, vascular volume, or use of oxygen or glucose [reviewed in (*44*)]. Even if the proliferation rate of cancer cells around normalized vessels was increased, this may well enhance therapeutic efficacy, because rapidly proliferating cells are more sensitive to radiation and to many cytotoxic drugs.

3) It is widely assumed that hypoxia leads to the death of cells. Therefore, alleviation of hypoxia during transient normalization of tumor vasculature should accelerate tumor growth. However, a growing body of evidence indicates that hypoxia may in fact promote cancer progression (*45*, *46*). These two competing effects of antiangiogenic therapy may cancel each other.

4) Finally, in some tumors, cancer cells depend on the same angiogenic growth factors (e.g., VEGF) for their survival as do the endothelial cells. In these tumors, antiangiogenic agents may kill both cancer cells and endothelial cells and will likely induce tumor regression—similar to hormone withdrawal from a hormone-dependent tumor (*47*)—despite vessel normalization.

For all these reasons, any acceleration in tumor growth during transient normalization is presumably masked by indirect and direct

killing of cancer cells by antiangiogenic agents. Thus, it is not surprising that tumor regression is slow and/or modest after antiangiogenic monotherapy despite a significant decrease in microvascular density (*31*).

Perspective

The approval of the first antiangiogenic agent for clinical use in patients with colorectal carcinoma has taught us many lessons, the most important of which is that these agents must be used in combination with agents that target cancer cells to have an appreciable impact on patient survival. Increasing the dose of antiangiogenic agent may harm normal tissues and destroy too much of the tumor vasculature, leading to hypoxia and poor drug delivery in the tumor and to toxicity in normal tissues. However, optimal doses and schedules of these reagents tailored to the angiogenic profile of tumors can normalize tumor vasculature and microenvironment without harming normal tissue.

At least three major challenges must be met before therapies based on this vascular normalization model can be successfully translated to the clinic. The first challenge is to determine which other direct or indirect antiangiogenic therapies lead to vascular normalization. In principle, any therapy that restores the balance between pro- and antiangiogenic molecules should induce normalization. Indeed, withdrawing hormones from a hormone-dependent tumor lowers VEGF levels and leads to vascular normalization (*47*). Recently, metronomic therapy—a drug delivery method in which low doses of chemotherapeutic agents are given at frequent intervals—has also been shown to increase the expression of thrombospondin-1, which is a potent endogenous angiogenesis inhibitor (*48*). Conceivably, this therapy might also induce normalization and improve oxygenation and drug penetration into tumors. Whether various synthetic kinase inhibitors (e.g., Novartis PTK787, Bayer 43-9006, Pfizer SU11248, and AstraZeneca AZD2171), endogenous inhibitors (e.g., angiostatin, endostatin, and tumstatin), antivasocrine agents [i.e., razoxane (*49*)], conventional chemotherapeutic agents [e.g., taxol (*50*)], and vascular targeting agents (*51–54*) do the same remains to be seen. Some of these agents may be effective because they target both stromal and cancer cells. To date, most clinical trials are designed primarily to measure changes in the size of the tumor and may therefore not shed light on changes in the vascular biology of tumors. Clinical studies, such as the rectal carcinoma study described earlier (*31*), and other ongoing translational clinical trials should help bridge the gaps in this aspect of our knowledge.

The second challenge is to identify suitable surrogate markers of changes in the structure and function of the tumor vasculature and to develop imaging technology that will help to identify the timing of the normalization window during antiangiogenesis therapy. Measurement of blood-vessel density requires tissue biopsy and provides little information on vessel function. Although imaging techniques are expensive and far from optimal, they can provide serial measures of vascular permeability, vascular volume, blood perfusion, and uptake of some drugs and can therefore be used to monitor the window of normalization in patients. The number of circulating mature endothelial cells and their less differentiated progenitors does decrease after VEGF blockade, both in animals and in patients, but whether this decline coincides with the normalization window is not known (*31*). During the course of therapy, serial blood measurements of molecules involved in vessel maturation have the potential to identify surrogate markers. PET with 18-fluoromisonidzole and MRI can provide some indication of tumor oxygenation (*55*, *56*) and might be useful for tracking the normalization window. Finally, the measurement of the interstitial fluid pressure is minimally invasive, inexpensive, and easy to implement for anatomically accessible tumors. Hence, this parameter could be used in the interim as a useful indicator of vessel function until novel noninvasive methods are developed.

The third challenge is to fill gaps in our understanding of the molecular and cellular mechanisms of the vascular normalization process (*29*). With rapid advances in genomic and proteomic technology and access to tumor tissues during the course of therapy, we can begin to monitor tumor response to antiangiogenic therapies at the molecular level.

Addressing all of these challenges not only will benefit patients with cancer but also may benefit patients with other diseases (*18*, *57*). For example, a low dose of anti-VEGF aptamer was recently shown to improve the vision of patients with macular degeneration, whereas a higher dose was ineffective (*58*). These principles may also be useful for regenerative medicine and tissue engineering, in which the goal is to create and maintain a functionally normal vasculature (*59*).

References and Notes

1. J. Folkman, *N. Engl. J. Med.* **285**, 1182 (1971).
2. J. C. Yang *et al.*, *N. Engl. J. Med.* **349**, 427 (2003).
3. M. A. Cobleigh *et al.*, *Semin. Oncol.* **30**, 117 (2003).
4. R. J. Mayer, *N. Engl. J. Med.* **350**, 2406 (2004).
5. H. Hurwitz *et al.*, *N. Engl. J. Med.* **350**, 2335 (2004).
6. B. A. Teicher, *Cancer Metastasis Rev.* **15**, 247 (1996).
7. D. J. Hicklin, L. M. Ellis, *J. Clin. Oncol.*, in press.
8. J. Ma *et al.*, *Cancer Res.* **61**, 5491 (2001).
9. R. Murata, Y. Nishimura, M. Hiraoka, *Int. J. Radiat. Oncol. Biol. Phys.* **37**, 1107 (1997).
10. B. M. Fenton, S. F. Paoni, I. Ding, *Radiother. Oncol.* **72**, 221 (2004).
11. R. K. Jain, *Nature Med.* **7**, 987 (2001).
12. P. Carmeliet, R. K. Jain, *Nature* **407**, 249 (2000).
13. R. K. Jain, in *Clinical Oncology*, M. Abeloff *et al.*, Eds. (Elsevier, New York, ed. 3, 2004), pp. 153–172.
14. T. P. Padera *et al.*, *Nature* **427**, 695 (2004).
15. D. P. Bottaro, L. A. Liotta, *Nature* **423**, 593 (2003).
16. S. K. Hobbs *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4607 (1998).
17. R. K. Jain, *Clin. Cancer Res.* **5**, 1605 (1999).
18. R. K. Jain, *Nature Med.* **9**, 685 (2003).
19. M. Ratner, *Nature Biotechnol.* **22**, 1198 (2004).
20. H. F. Dvorak, *J. Clin. Oncol.* **20**, 4368 (2002).
21. N. Ferrara, K. J. Hillan, H. P. Gerber, W. Novotny, *Nature Rev. Drug Discov.* **3**, 391 (2004).
22. J. A. Nagy *et al.*, *J. Exp. Med.* **196**, 1497 (2002).
23. C. G. Lee *et al.*, *Cancer Res.* **60**, 5565 (2000).
24. R. T. Tong *et al.*, *Cancer Res.* **64**, 3731 (2004).
25. F. Yuan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14765 (1996).
26. A. Kadambi *et al.*, *Cancer Res.* **61**, 2404 (2001).
27. T. Inai *et al.*, *Am. J. Pathol.* **165**, 35 (2004).
28. H. Wildiers *et al.*, *Br. J. Cancer* **88**, 1979 (2003).
29. F. Winkler *et al.*, *Cancer Cell* **6**, 553 (2004).
30. Y. Gazit *et al.*, *Microcirculation* **4**, 395 (1997).
31. C. G. Willett *et al.*, *Nature Med.* **10**, 145 (2004).
32. B. Morgan *et al.*, *J. Clin. Oncol.* **21**, 3955 (2003).
33. H. Q. Xiong *et al.*, *Invest. New Drugs* **22**, 459 (2004).
34. R. S. Herbst *et al.*, *J. Clin. Oncol.* **20**, 3804 (2002).
35. F. Baffert *et al.*, *Circ. Res.* **94**, 984 (2004).
36. M. Relf *et al.*, *Cancer Res.* **57**, 963 (1997).
37. Y. Izumi, L. Xu, E. di Tomaso, D. Fukumura, R. K. Jain, *Nature* **416**, 279 (2002).
38. B. Goldman, *J. Natl. Cancer Inst.* **95**, 1744 (2003).
39. P. Perrotte *et al.*, *Clin. Cancer Res.* **5**, 257 (1999).
40. S. J. Kim *et al.*, *Cancer Res.* **64**, 4201 (2004).
41. R. Kerbel, J. Folkman, *Nature Rev. Cancer* **2**, 727 (2002).
42. B. Vogelstein, K. W. Kinzler, *Nature Med.* **10**, 789 (2004).
43. D. Hanahan, R. A. Weinberg, *Cell* **100**, 57 (2000).
44. P. M. Gullino, in *Biomedical Thermology*, M. Gautherie, E. Albert, Eds. (A. R. Liss, New York, 1982), pp. 1–20.
45. L. Flintoft, *Nature Rev. Cancer* **4**, 754 (2004).
46. D. A. Nelson *et al.*, *Genes Dev.* **18**, 2095 (2004).
47. R. K. Jain *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10820 (1998).
48. R. S. Kerbel, B. A. Kamen, *Nature Rev. Cancer* **4**, 423 (2004).
49. K. Hellmann, *Clin. Exp. Metastasis* **20**, 95 (2003).
50. G. Griffon-Etienne, Y. Boucher, C. Brekken, H. D. Suit, R. K. Jain, *Cancer Res.* **59**, 3776 (1999).
51. D. W. Siemann, D. J. Chaplin, M. R. Horsman, *Cancer* **100**, 2491 (2004).
52. P. E. Thorpe, *Clin. Cancer Res.* **10**, 415 (2004).
53. E. Ruoslahti, *Nature Rev. Cancer* **2**, 83 (2002).
54. Vascular targeting agents (VTAs) and cytotoxic agents presumably target different regions of a tumor. VTAs indirectly kill cancer cells by choking the tumor's blood supply, but leave a viable rim of cancer cells. It is the latter cells that are the target of chemotherapy or radiation therapy. The order in which the tumor and vascular cells are targeted may be key to the success of this approach. The greatest therapeutic benefit may come when blood vessels are destroyed after chemotherapy. In this scenario, the cytotoxic drug will be retained in the tumor for a longer period because there is no longer an efficient means of egress. It is worth noting here that the efficacy of these agents has not yet been proven in the clinic with phase III trials. Even in preclinical studies, the effects of combined VTAs and conventional therapies have been additive at best, whereas the effects of combined antibody to VEGFR2 and fractionated radiation—given during the normalization window—have been synergistic (i.e., more than additive) (*29*).
55. J. S. Rasey *et al.*, *Int. J. Radiat. Oncol. Biol. Phys.* **36**, 417 (1996).
56. S. Gross, A. Gilead, A. Scherz, M. Neeman, Y. Salomon, *Nature Med.* **9**, 1327 (2003).
57. P. Carmeliet, *Nature Med.* **9**, 653 (2003).
58. FDA, www.thepinksheet.com/nr/FDC/SupportingDocs/pink/2004/ 040830_Macugen_fda_materials.pdf (2004).
59. N. Koike *et al.*, *Nature* **428**, 138 (2004).

60. The author thanks the members of the Steele Lab, especially M. Booth, Y. Boucher, E. diTomaso, D. G. Duda, D. Fukumura, S. Kozin, L. Munn, T. Padera, R. Tong, C. Willett, and F. Winkler; and B. Chabner, H. Chen, L. Ellis, J. Folkman, R. Jones, K. Novak, J. Samson, R. Weinberg, and B. Zetter for their helpful comments on this manuscript. The author is a consultant to AstraZeneca. This work was supported by grants from the National Cancer Institute.

10.1126/science.1104819

INTRODUCTION

Everywhere You Turn

Wherever matter concentrates in the vast vacuum of space, chances are you will find a disk. Simple physics makes it so; all you need is a spinning sphere of matter collapsing under its own gravitational pull. As the core contracts, it rotates faster to conserve angular momentum. This spin causes the matter around the core to fall onto the equatorial plane, forming a flattened disk of gas and solids. Angular momentum creeps outward, keeping the core from breaking apart. Meanwhile, disk material moves inward along the equatorial plane, eventually either feeding the core or forming other orbiting objects around the core. That oft-repeated scenario makes disks essential building blocks of the universe. This special section covers the many different flavors of disks in space and what they can tell us about the formation of everything from giant gas planets to galaxies.

The solar system got its start when a molecular cloud collapsed to form the Sun and a circumstellar disk of gas and dust. As Connolly notes in his Viewpoint (p. 75), chondritic meteorites contain primordial dust from other nearby stars, evidence that the Sun formed within a cluster of stars. In his Review, Gladman (p. 71) takes us to the edge of the solar system, where remnants of the circumstellar disk are dispersed in the Kuiper belt. The belt is more extensive and more structured than previously thought, and its structure holds clues to planetary formation, planetary migration, possible rogue planets, and the close passage of other stars.

Over the past decade, as Greaves describes in her Review (p. 68), the search for circumstellar disks with possible planets around other stars has intensified. Planets around Sunlike stars are now considered to be ubiquitous. Current models of planet formation require a disk full of gas and dust to swirl and sway around the star long enough to accrete giant gas planets. Observations of the different stages of the evolution of circumstellar disks are helping to refine these models.

For decades, astrophysicists thought that disk-shaped spiral galaxies turn into featureless balls of stars when they collide with other galaxies. In the first of his two News stories (p. 64), however, Irion describes evidence that the spirals (which include our own Milky Way galaxy) can be surprisingly resilient. Meanwhile, other astronomers are probing how disks of matter pulled from a companion star trigger the nuclear explosions that turn white dwarf stars into type Ia supernovas (p. 66): cosmic flares that help gauge the expansion of the universe.

Narayan and Quataert (p. 77) end this special issue with a Review of the most illuminating evidence for black holes: the accretion disks that surround them. As a black hole accretes gas, the gas radiates and provides a thermal signature of the mass, rate of spin, and location of the event horizon of the black hole.

This quick tour of disks in space highlights their simplicity and ubiquity. As modeling and observations continue to provide more details of disk complexity, their utility for resolving fundamental mysteries of space, such as planet formation and black hole jets, will grow.

–LINDA ROWAN, DANIEL CLERY, ROBERT COONTZ

CONTENTS

CREDIT: GEMINI OBSERVATORY/GMOSTEAM

Science

Published by AAAS

NEWS

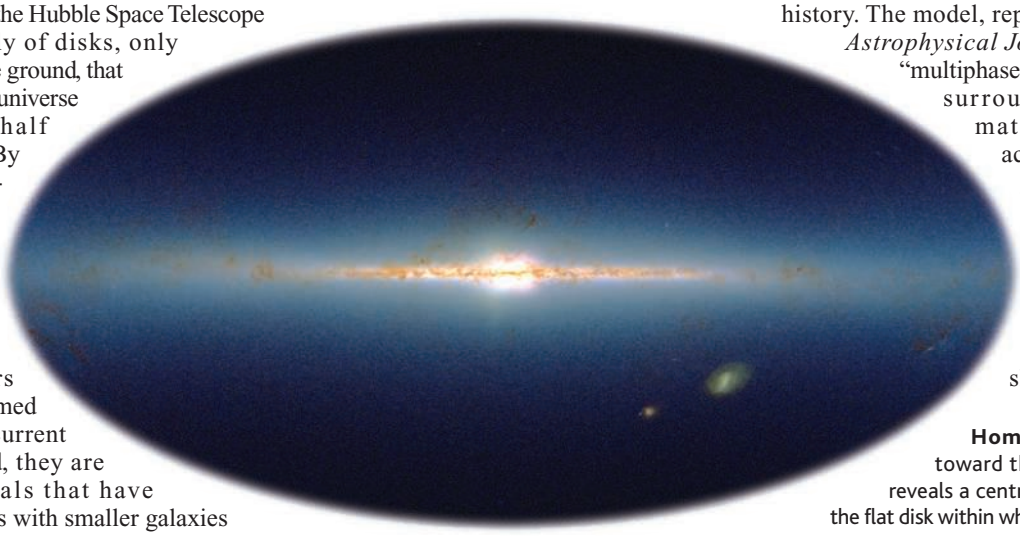
As the Galaxies Turn

Spiral disk galaxies, serene icons of the universe, are hardy survivors of a battering cosmic history

Gravity conspires to produce two dominant shapes in astronomy: spheres and disks. Both are on display in spiral galaxies, home to perhaps half the stars in the universe. Spherical central bulges of old yellow suns glow serenely, girdled by a disk consisting of curved arms of hot new stars and dark bands of dust. Such grand stellar disks, long the pinups of astronomy buffs, now play a starring role in studies of how galaxies have evolved.

Surveys with the Hubble Space Telescope reveal a panoply of disks, only hinted at from the ground, that existed when the universe was less than half its current age. By dating and classifying this huge population, astronomers are recognizing that spiral galaxies are not delicate flowers that have blossomed slowly to their current display. Instead, they are tough perennials that have survived mergers with smaller galaxies and—on occasion—crushing collisions with big ones throughout billions of years of cosmic time.

In our edge-on view of the Milky Way's plane, we gaze upon just such a stalwart bisecting the night, one that undoubtedly consumed other galaxies. The Milky Way's disk provides clues to this history, but the sleuthing is tough because we're embedded within it. "We have an opportunity to understand it at a much deeper level than other galaxies, because we can measure the motions of individual stars," says astronomer Heidi Jo Newberg of Rensselaer Polytechnic Institute in Troy, New York. "But we're really just starting."



Home. An infrared view toward the Milky Way's core reveals a central bulge of stars and the flat disk within which we live.

CREDIT: IPAC/CALTECH/2MASS/NASA

It's all in the gas

The disks we see today took a long time to develop. "Almost all star formation was in clumps and chaotic structures" for roughly the first 4 billion years of cosmic history, says astronomer Sidney van den Bergh of the Dominion Astrophysical Observatory in Victoria, British Columbia. But during the next 1 billion to 2 billion years, recognizable features started to form under the inexorable pull of gravity.

Astronomers believe that a typical primitive galaxy was a bloated cloud, slowly rotating and rich in warm gas that had not yet coalesced into many stars. Energy escaped from the cloud as atoms and molecules collided and radiated light. Gravity pulled the cooling gas more tightly together, forcing more frequent collisions, but it would have kept its original angular momentum. As time marched on, the fledgling galaxy flattened and spun faster and faster.

"The final state of a runaway collapse is a thin disk where all particles go in exactly circular orbits," says astrophysicist Julio Navarro of the University of Victoria, British Columbia. But a galaxy isn't an idealized whorl of gas, he notes: "When the gas collects into tiny little packets of stars, you get a collection of bullets that never collide." Without energy-robbing collisions, a star-filled disk cannot settle down if it gets perturbed by another young galaxy plunging into it—a common event in the cosmic past. Instead, stars tend to scatter into spherical swarms, like a disturbed hive of bees.

This is exactly what happened when astronomers constructed computer simulations of evolving galaxies dominated by stars. "Disks are very fragile, dynamical entities. Mergers mess them up," Navarro says. But if mergers and collisions were so common in the early universe, why don't we see the sky full of formless elliptical galaxies?

The influence of gas is the key, Navarro and others now agree. Effervescent gas would have damped out the otherwise shattering effects of major mergers. Adolescent galaxies could have kept gas stirred up in plenty of ways: intense ultraviolet light from massive newborn stars, shock waves from supernova explosions, or outpourings of energy from vigorous cores.

Recent simulations have shown this damping effect of gas in action. For instance, a team led by graduate student Brant Robertson of Harvard University in Cambridge, Massachusetts, produced one of the first realistic disk galaxies in a simulation that spans cosmic history. The model, reported in the 1 May *Astrophysical Journal*, relies on a "multiphase gas" of cold clouds surrounded by hotter material, which more accurately captures a galaxy's interstellar environment. This hybrid recipe preserves gas during mergers and stabilizes the disk against external onslaughts, Robertson says. The approach works, but it's only a start: Just 1 of 20 simulated galaxies ended up with a flat pinwheel of stars and gas, compared with about half in the real universe. Improved models may need to churn up the gas even more with early bouts of star formation, other researchers believe.

And in new work submitted to *Astrophysical Journal Letters*, two of Robertson's co-authors demonstrate that a classic spiral galaxy can emerge even from the wreckage of a violent collision. Astrophysicists Volker Springel of the Max Planck Institute for Astrophysics in Garching, Germany, and Lars Hernquist of Harvard plowed two simulated gas-rich disks into each other. The concussive impact sparked a blaze of star birth, but enough gas remained to settle the merged object into a flat superdisk with clear spiral arms. "If disks can 'survive' even major mergers, they are probably less fragile than previously thought," the researchers write.

Forty thousand personalities

Simulations are an alluring way to peer back to galactic youth, but nothing beats the real thing.

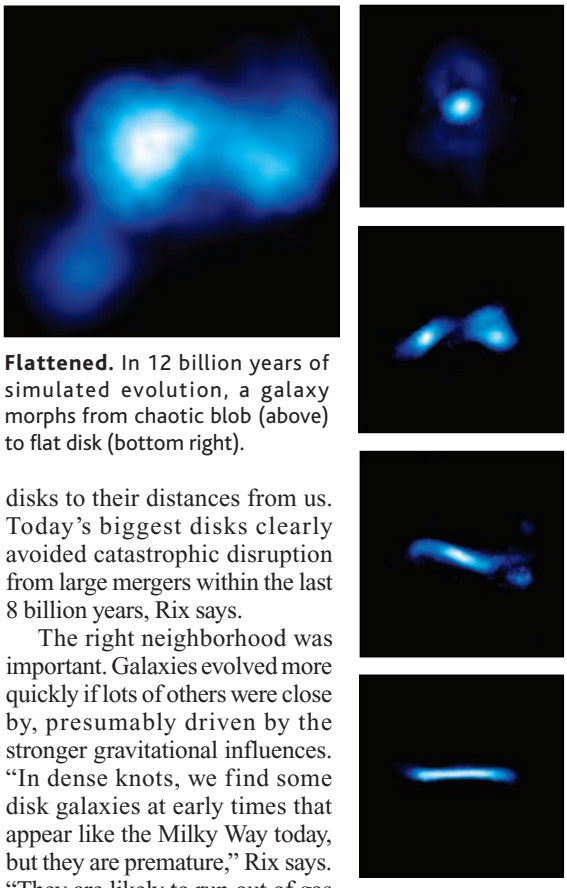
Published by AAAS

Enter GEMS—Galaxy Evolution From Morphology and Spectral Energy Distributions—an ambitious program to deduce how overall populations of galaxies have evolved. GEMS studies about 40,000 galaxies in the Hubble Space Telescope's largest contiguous color image of the sky: 150 times as broad as its "deep field" image taken in 1994. Astronomers have good distance estimates to some 10,000 of those galaxies, from spectra obtained at the European Southern Observatory's 2.2-meter telescope at La Silla, Chile.

For astronomers, GEMS has been as transforming as seeing a photo album of hundreds of ancestors rather than just a few faded snapshots. "From the ground, these galaxies are dots. But from Hubble, each one gets a personality," says lead scientist Hans-Walter Rix of the Max Planck Institute for Astronomy in Heidelberg, Germany.

After more than a year, the GEMS team can make firm statements about the life and times of disks since the universe was about 5 billion years old. For example, the team charted the hottest starlight from newborn suns. "For the last 8 billion years, by far the largest majority of stars have formed in disk galaxies that start to resemble our Milky Way," Rix says. In contrast, elliptical galaxies had their heyday of spawning stars billions of years earlier.

At the outer reaches of its survey, the team sees what Rix calls "a sufficient number of galaxies with a bulge in the middle and small disks around them." These objects, he says, are most likely the ancestors of large disk galaxies such as the Milky Way and nearby Andromeda. Moreover, such galaxies grew their disks from the inside out, a maturation that the team traces by comparing the sizes of disks to their distances from us. Today's biggest disks clearly avoided catastrophic disruption from large mergers within the last 8 billion years, Rix says.



Flattened. In 12 billion years of simulated evolution, a galaxy morphs from chaotic blob (above) to flat disk (bottom right).

The right neighborhood was important. Galaxies evolved more quickly if lots of others were close by, presumably driven by the stronger gravitational influences. "In dense knots, we find some disk galaxies at early times that appear like the Milky Way today, but they are premature," Rix says. "They are likely to run out of gas and star formation, merge, and become ellipticals. That is their fate."

Step up to the bar

Not all is symmetrical in the realm of spiral galaxies. About two-thirds of all disks sport "bars"—elongated concentrations of stars embracing the galactic cores. Our Milky Way has one: a stubby bar first suspected in the 1980s and recently mapped by laborious census of a distinctive class of stars within the disk.

Bars can alter disk galaxies by redistributing mass and angular momentum. "Any kind of perturbation in a cold disk tends to form bars or spiral arms," says astronomer Shardha Jogee of the University of Texas, Austin. Once formed, a bar tugs gravitationally on gas and pulls it toward the center of the galaxy, triggering the birth of new stars. In theory, this may sow the seeds of a bar's destruction. Some early simulations showed that a central buildup of mass propels passing stars farther out onto great looping paths, dissolving the bar and its narrowly confined stellar orbits.

But more recently, astronomers have wondered how quickly these transitions might happen. "The evolution from barred to unbarred and back again can go on in the lifetime of a galaxy, but there has always been a lot of question about how fast this process is," says astronomer Mousumi Das of the University of Maryland, College Park. GEMS points to a slower transformation than expected. The team, led by Jogee, found a constant ratio of strongly barred to unbarred galaxies at all epochs. The structures survive at least 2 billion years, if not much longer, the authors concluded in the 10 November *Astrophysical Journal Letters*.

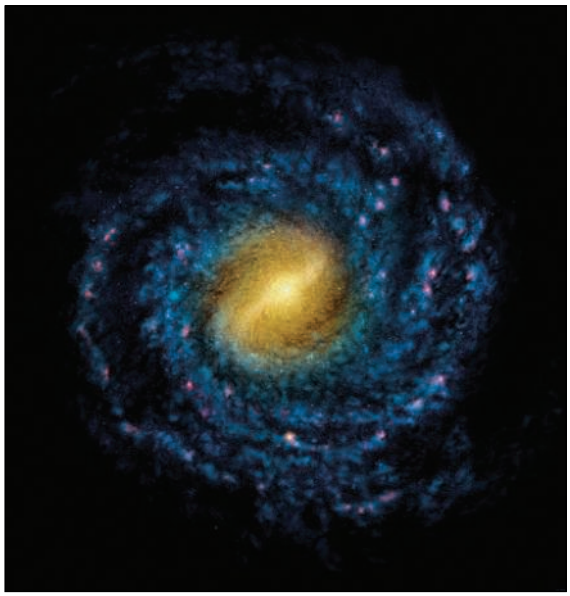
Another valuable tracer of a galaxy's history is its so-called thick disk, a smattering of older stars that wander above and below the main disk. Astronomers aren't yet sure how stars in the Milky Way's thick disk got there. In one popular scenario, a galaxy merger harassed the stars out of their cozy orbits in the thin disk, perhaps 10 billion years ago. Because there are no stars younger than that in the thick disk, that event probably was the galaxy's last noteworthy consolidation, says astronomer Rosemary Wyse of Johns Hopkins University in Baltimore, Maryland.

However, Julio Navarro and his colleagues think they see imprints of a more fascinating tale. Scrutiny of the motions and chemical compositions of stars in the thick disk reveal a few odd groupings that have properties dissimilar to those of the rest of the galaxy. The team proposes, provocatively, that the thick disk is not a puffed-up set of the Milky Way's own stars but is shot through with aliens. Arcturus, a bright star not far from the sun, could be one such immigrant from a long-ago devoured galaxy.

The next step for astronomers involved in this galactic archaeology will be a thorough charting of the motions of millions of Milky Way stars all around us. One such effort, the Radial Velocity Experiment, is under way at the 1.2-meter U.K. Schmidt Telescope in Siding Spring, Australia. And a proposed extension of the U.S.-led Sloan Digital Sky Survey would examine stars in the galaxy's crowded plane, a region the survey has largely avoided.

Starting in 2011, the European Space Agency's Gaia satellite will scrutinize a billion stars, fully 1% of the galaxy's population. We may then learn how our familiar disk has kept itself together in a universe full of disorderly influences.

–ROBERT IRION



Looking back. A journey far above the Milky Way's disk might reveal this view of its spiral arms and central bar.

CREDITS: (SERIES AT TOP) B. ROBERTSON AND L. HERNQUIST/HARVARD, N. YOSHIDA/UNIV. OF NAGOYA, AND V. SPRINGEL/MPA; (BOTTOM) CHRIS BUTLER

Published by AAAS

NEWS

Disks of Destruction

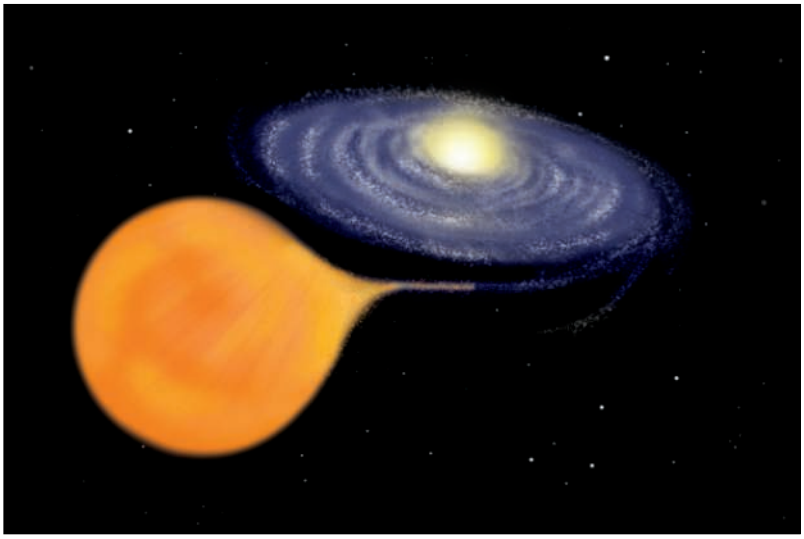
Exploding white dwarfs are a key yardstick of the cosmos, but how does gas spiraling onto their surfaces make these stellar corpses blow up?

Most white dwarfs live gently and disappear silently. The remains of stars like our sun, white dwarfs usually cool down for billions of years and fade into black cinders. But some of these stars refuse to go quietly. Aided by matter stolen from other stars, they explode in planet-sized fusion bombs that can outshine entire galaxies.

These blasts, called type Ia supernovas, became famous in 1998 when astronomers used them as flashbulbs of standard brightness to show that the cosmos is expanding at an accelerating rate. Newer studies have validated that startling claim (*Science*, 19 December 2003, p. 2038). Today, the supernovas are central in efforts to decipher "dark energy," the bizarre force driving the universe's hastening growth.

Yet tough puzzles confound the cottage industry that now studies type Ia supernovas. Astronomers don't yet know how a white dwarf gains enough mass to seal its doom. In the most popular explanation, the dwarf slowly strips gas from a nearby swollen companion star approaching the end of its life; the stolen gas forms an accretion disk that ultimately sparks the white dwarf's spectacular demise. An alternative explanation sees the donor as another white dwarf spiraling toward a messy crash. What's more, theorists disagree deeply about how a dwarf actually blows up.

With astronomers planning to expand their catalog of type Ia supernovas by thousands and increase their use as cosmological measures, these questions are unsettling. "If you don't understand type Ias, do you trust using them for cosmic evolution?" asks astrophysicist David Branch of the University of Oklahoma, Norman. "Everyone would feel better if we had some understanding of the tools we're using."



Costly meal. If a white dwarf (upper right) steals enough gas from a companion star via an accretion disk, the dwarf may explode in a type Ia supernova.

ILLUSTRATION: J. MOGUL/*SCIENCE*

Shattered to smithereens

Ordinary low-mass stars create white dwarfs when their central nuclear fires burn out. Gas in the core gets crushed to an Earth-sized ball until electrons—pushed to ever-higher orbital energies—resist further squeezing. But if by some method a greedy dwarf attracts extra gas and its mass approaches the magic "Chandrasekhar limit" of 1.44 times the mass of our sun, its peaceful retirement is over.

Researchers at least see eye to eye on the basics of what happens next. "Everyone agrees that type Ia supernovas are thermonuclear disruptions of mass-accreting white dwarfs," says astronomer Mario Livio of the Space Telescope Science Institute (STScI) in Baltimore, Maryland. Close to the critical mass, the increased pressure ignites a chain reaction in the core. New fusion rips the dwarf apart within seconds, Livio says: "The whole thing is shattered to smithereens. There is nothing left behind."

The debris propelled out by each blast yields the only visible clues. Telescopes detect the spectral signatures of the heavy elements iron and nickel, as well as silicon, calcium, and magnesium. The patterns—especially the amount of radioactive nickel, which powers the supernova's light display—match what astrophysicists expect from the sudden combustion of a white dwarf made mostly of carbon and oxygen.

The first puzzle for astronomers is the origin of those fatal dollops of extra matter. The favored source is a full-fledged star: a binary companion to the white dwarf. When this star's core starts to run out of nuclear fuel, it burns hotter and puffs up its outer atmosphere. If the white dwarf orbits closely enough, it can swipe this loosely bound gas. Gravity pulls the gas into an accretion disk that dribbles matter onto the white dwarf's surface, and the stellar thief heats up with new life.

But the transfer must happen at just the right rate: roughly one 10-millionth of a solar mass per year, or 1/30 of Earth's mass. Any higher or lower than this, theory suggests, and thermonuclear flashes on the white dwarf's surface or vigorous winds can expel much of the matter gained from the accretion disk. This Goldilocks requirement sounds frightfully specific, but astronomers think they see white dwarfs accreting in binary systems at something like the right rate. "There appear to be enough of these to get the right kind of statistics for type Ia supernovas, roughly 3 per 1000 years in the Milky Way," Livio says.

Hot on the trail

Recent research has bolstered this scenario. Astronomers identified a star they believe supplied the matter to trigger the most recent type Ia event known in our galaxy: Tycho Brahe's supernova from 1572. A team led by Pilar Ruiz-Lapuente of the University of Barcelona, Spain, found the star dashing through the sizzling supernova remnant three times faster than neighboring stars—just as if it had been set free from a high-velocity binary orbit. In the 28 October issue of *Nature*, the team reported that the star looks like an aged version of our sun but puffed up, as if by a recent blast wave.

Another study took a different tack to reach a similar conclusion about donor stars. STScI astronomer Louis-Gregory Strolger and colleagues examined 42 distant type Ia supernovas, between 2.4 billion and 9.5 billion light-years away. The team compared how many explosions popped off in different epochs of the universe's history with the times that stars formed. In the 20 September *Astrophysical Journal*, the astronomers concluded that there is an average time delay of 4 billion years for new

Published by AAAS

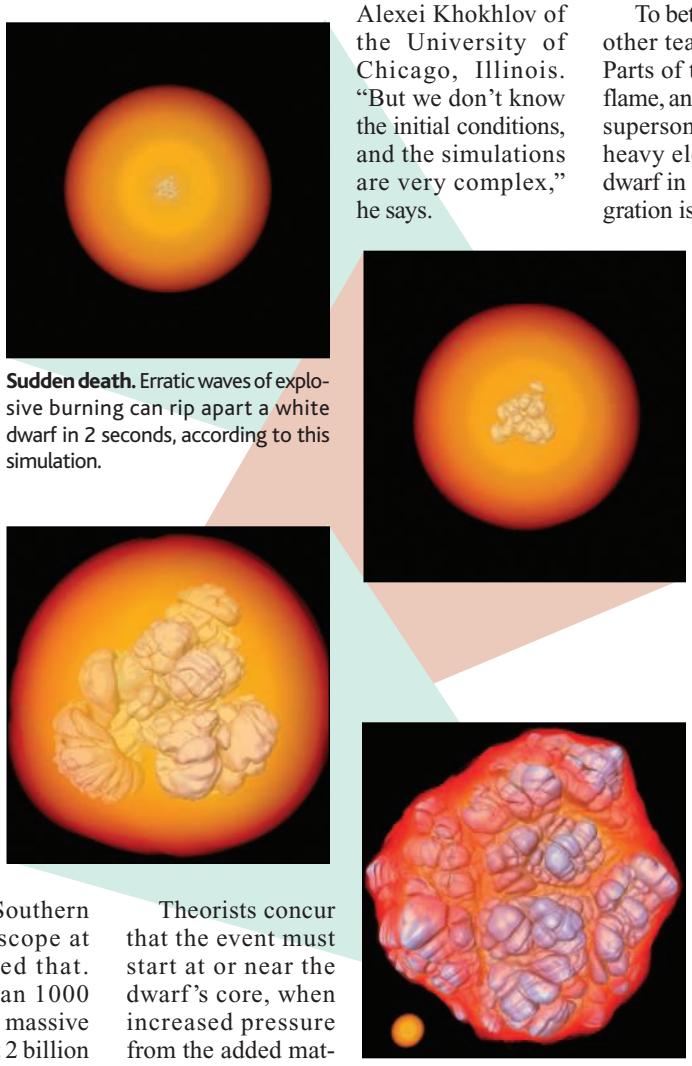
star systems to start spawning type Ia supernovas. That's about right for one member of a binary pair to expand at the end of adulthood and lose gas to a compact companion, Strolger says.

But one very different idea lingers stubbornly. Some astronomers hold that type Ia supernovas result from the ruinous mergers of two white dwarfs. Longtime proponent Icko Iben of the University of Illinois, Urbana-Champaign, points to differences from one supernova to the next as a natural outcome of mergers between dwarfs of varying masses. Moreover, type Ia supernovas in ancient elliptical galaxies—where no new stars have formed for billions of years—are best explained by binary pairs of old dwarfs that gradually spiral together, Iben believes.

Most damning of all, in Iben's view, is the missing hydrogen. Bloated stars that donate mass to a white dwarf should pollute the whole environment with hydrogen, he says. And yet astronomers have seen evidence of hydrogen in just one type Ia supernova, in 2002. It's totally lacking in the rest, but crashes of hydrogen-free white dwarfs explain that logically. "To me that is a very powerful argument," Iben says.

For years, people dismissed this idea by noting that no suitable merger candidates were seen in the Milky Way. But a recent survey by the European Southern Observatory's Very Large Telescope at Cerro Paranal in Chile changed that. Astronomers observed more than 1000 white dwarfs and discovered one massive binary pair that will merge in about 2 billion years, says survey leader Ralf Napiwotzki of the University of Leicester, U.K.

One prominent team is dubious that such pairs will explode. Astrophysicists Ken'ichi Nomoto of the University of Tokyo and Hideyuki Saio of Tohoku University in Japan verified earlier calculations that the lighter of two merging white dwarfs would break up completely and form a thick accretion disk around the more massive one. As this material streams onto the dominant dwarf, carbon within the heated star burns into oxygen, magnesium, and neon, Nomoto says. Once the composition changes, a sudden flash of fusion can no longer occur, and the dwarf collapses "peacefully" into a neutron star. But other theorists maintain that synchronized rotation between the two dwarfs would force a different style of accretion, triggering an all-consuming explosion.



Sudden death. Erratic waves of explosive burning can rip apart a white dwarf in 2 seconds, according to this simulation.

CREDITS: F. RÖPKE AND W. HILLEBRANDT/MPIA

Got a match?

You might think it would be easy to explain how a single white dwarf blows up, but the story is anything but neat. Theorists know the nuclear physics well, says astrophysicist Alexei Khokhlov of the University of Chicago, Illinois. "But we don't know the initial conditions, and the simulations are very complex," he says.

Theorists concur that the event must start at or near the dwarf's core, when increased pressure from the added matter sparks runaway fusion among carbon and oxygen atoms. But even this initial step causes angst. "The most crucial question in these detailed models is how the flame ignites, but there's a lot of hand waving involved," says astrophysicist Wolfgang Hillebrandt of the Max Planck Institute for Astrophysics in Garching, Germany.

Hillebrandt's colleague Friedrich Röpke envisions "a central ignition with a foamy structure. One expects strong convection before the ignition, so it should ignite in several spots distributed around the center." This wave of burning, called deflagration, spreads erratically through the star's interior at subsonic speeds as the material convects, leaving unburned pockets of carbon and oxygen.

At this point, theorists diverge. The German group thinks deflagration can explain the entire supernova, blowing it apart in about 2 seconds. But others point out that such an explosion would produce debris that is too rich in unburned light elements and too "mixed" from the inside out. Spectra from real type Ia supernovas suggest a more layered explosion, says Oklahoma's Branch.

To better match those data, simulations by other teams invoke "delayed detonation." Parts of the dwarf burn with a deflagration flame, and then the whole thing explodes with supersonic shock waves. The shocks forge heavy elements throughout the rest of the dwarf in tenths of a second. Indeed, if deflagration is like a wind-driven fire that torches some trees and skips others, detonation is a bomb that incinerates the whole forest. The problem is that no one knows how to set the bomb off. "There is no known physical mechanism to convert deflagration to a supersonic detonation," says Röpke.

One possible solution recently startled the field. Theorists led by Tomasz Plewa of the University of Chicago set off a simulated detonation with a supersonic bubble that rose from the dwarf's interior, rather like a jellyfish. When the bubble broke the surface, it unleashed waves that raced around the white dwarf—confined by gravity—and collided on the far side. The smashup was violent enough to incinerate the star in an asymmetric blast. No one quite knows what to make of the weird sequence of events, published in the 1 September *Astrophysical Journal Letters*.

Observers watch these efforts with anticipation and frustration. "We need a guide from theorists to understand the differences in these objects, and I'm not sure whether theory is up to it," says astronomer Adam Riess of STScI. At one recent talk, Riess says, he counted 15 "knobs" that the theorist turned to adjust his model. "That's a huge number of ways to produce a supernova," he says. "I found that a little distressing."

Keep the faith, says Khokhlov: "We have great hope that the explosion mechanism is somehow channeled through a very narrow evolutionary bottleneck," with a set of unique solutions that theorists will unveil. Many hope that will happen soon, because without it there is an element of doubt when astronomers claim to divine the history and fate of the entire universe.

—ROBERT IRION

Published by AAAS

REVIEW

Disks Around Stars and the Growth of Planetary Systems

Jane S. Greaves

Circumstellar disks play a vital evolutionary role, providing a way to move gas inward and onto a young star. The outward transfer of angular momentum allows the star to contract without breaking up, and the remnant disk of gas and particles is the reservoir for forming planets. High-resolution spectroscopy is uncovering planetary dynamics and motion within the remnant disk, and imaging at infrared to millimeter wavelengths resolves disk structure over billions of years of evolution. Most stars are born with a disk, and models of planet formation need to form such bodies from the disk material within the disk's 10-million-year life-span.

Disks around stars are flattened concentrations of gas and particles that follow orbits much like those of planets. A star forms from the collapse of a self-gravitating core within a molecular gas cloud, which generally has some rotation. Conservation of angular momentum implies that as the core contracts, its rotation must speed up, and this prevents the particles from falling in a direct line onto the growing star. Apart from a few particles along the orbital poles that can fall in because they have no rotation, all the others must first settle down onto an equatorial belt; it is this process that results in a circumstellar disk.

Disks are ubiquitous in astronomy because conservation of angular momentum is a universal law and almost every astronomical object starts with some rotation, hence contracting toward a plane rather than a point. Thus, disks occur frequently and on a wide range of scales, from the ensembles of billions of stars forming a flat spiral galaxy down to tiny energetic accretion zones that surround black holes (*1*). Disks around stars were recognized as a likely possibility long before they could be imaged. Recognition that the orbits of the Sun's planets were all close to one plane led to ideas that they could have formed out of a flat disk. In the modern era, nebulosity began to be seen around stars recognized as young from their high luminosities (i.e., still contracting and so having larger emitting surfaces), and in some cases central disks were suspected. In the 1960s, excess infrared (IR) emission was detected toward some young stars, leading to suggestions that particles absorb optical stellar light and reradiate it at longer wavelengths. Astronomers realized (*2*) that disks around young stars could both produce a warm excess of IR emission and regulate the stellar angular momentum. By the 1970s, disks were being used as explanations for polarized light scattered in the vicinity of stars, orbital motion of gas concentrations with maser emission, and channels for narrow stellar jets (*3*–*5*). However, it was only in the 1980s that true disk images were obtained; the first really compelling one was the star β Pictoris (*6*). This thin disk of light-scattering particles, seen edge-on from Earth, brought into prominence the idea that planetary cores could form by grain coagulation in disks. Some recent examples of disks in diverse star-forming clusters are shown in Fig. 1.

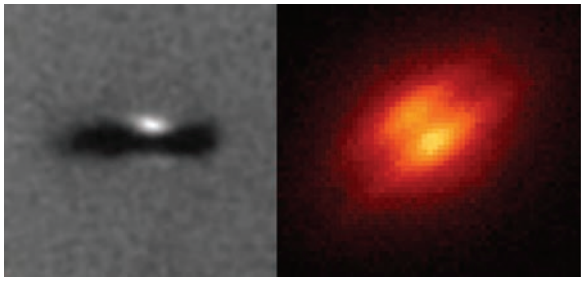


Fig. 1. Two circumstellar disks seen edge-on around stars with an age of about 1 Myear. (**Left**) Disk in the dense Orion nebula cluster (*45*); image is courtesy of NASA and Space Telescope Science Institute. (**Right**) Disk in the loose MBM 12 association (*46*); image courtesy Gemini Observatory–Association of Universities for Research in Astronomy. The light from the star is scattered toward us by disk particles, and in Orion the disk is seen in silhouette against the light from the background nebula. The dark lanes occur in the disk midplane where direct light from the star is blocked by dust.

In retrospect, it seems obvious that the material to make a star cannot be spherically spread out, or we would not be able to see the light of the star itself. A flattened disk allows light out along our line of sight, so only disks seen edge-on block the starlight. However, the term "shell" instead of "disk" appeared in the literature into the 1980s, and it should be remembered that, when "disk" is used now, this may be an assumption when no image exists. It is not certain that all stars are born from and within disks, although it is hard to explain their formation otherwise. More exotic processes could occur, such as the merger of two protostars (*7*). However, observations of young clusters of stars find over 80% have IR excesses (*8*), suggesting that evolution within a disk is typical.

For simplicity, subsequent discussion of disk evolution is focused on Sun-like stars. Their circumstellar disks contain particles totalling as much as 15% of the stellar mass, orbiting at modest Keplerian speeds (*9*) of 30 down to 3 km/s at 1 to 100 AU from a 1 $M_{\odot}$ star, where 1 astronomical unit (AU) is the Sun-Earth distance. Disk radii range from the size of Saturn's orbit (just under 10 AU) up to 800 AU in loose associations of stars (*10*). In some cases, the disk may be circumbinary, encircling two stars rather than one [for example, GG Tau Aa and Ab (*10*)]. These systems are poorly studied but important, because about two-thirds of stars like the Sun have such stellar companions (*11*).

Attempts have been made to classify the phases of evolution of a disk by stellar age (Fig. 2 and Table 1). For the first 10 Myear, the disk is often described as protoplanetary, possessing a large reservoir of dust and gas. This era is divided into subclasses 0 to III on the basis of the shapes of the optical and near-IR spectra of the star-disk system. These classes should relate to the relative masses of the star and the circumstellar material in a progression reflecting growth and then clearing of the disk. Consequently there is more direct and less reradiated stellar light at late times. The classifications do not represent an exact evolutionary sequence, because the disk mass distribution and angle from which we view it affect whether radiation emerges or is blocked. Consequently, stars in different observational subclasses may have similar ages. (Globally, they are often referred to as T Tauri stars after the prototype T Tau, which is now considered a class II object.) Beyond this 10-Myear period, the disk mass drops abruptly by a large factor, with rare examples of transition disks detected, such as β Pictoris (Fig. 2 and

Physics and Astronomy, University of St. Andrews, North Haugh, St. Andrews, Fife KY16 9SS, UK.

Table 1). After this is a much longer phase where dust particles are regenerated by collisions of comets and asteroids made in the protoplanetary era. This secondary dust makes up a debris disk. The Sun has only a sparse ring of such comets, the Kuiper Belt (*12*), in which collisions are infrequent and so little dust is generated. The dust of the outer Solar System has been inferred from impacts on spacecraft such as Pioneer (*13*); radiation from the Kuiper Belt is hard to detect from our position inside it.

Age also affects the composition of circumstellar disks. At early times, protoplanetary disks are dominated by hydrogen gas, plus traces of other common atoms such as carbon, nitrogen, and oxygen and simple molecules such as carbon monoxide. This gas coexists with solid particles made of carbon, silicates, and frozen ices, collectively referred to as dust grains (*14*). Debris disks contain larger grains because they are pieces broken off of colliding bodies. It might be expected that small transient gas clouds would appear because ices are vaporized in the impacts, but no gas has yet been detected (*15*).

The role of the circumstellar disk changes over the life of the star. A class 0 protostar collects 90% of its final mass from the cloud core in the first 0.1 Myear (*16*); nearly all of this mass must be processed through the disk because it cannot fall directly onto the star. The disk may then help to keep the star from breaking apart. As the protostar contracts into a small, dense body capable of nuclear fusion, the rotation increases to conserve angular momentum, and this rotation speed can become so great that the star can break up. A disk can act as a stellar brake that transports more than 99% of the angular momentum outward (*17*), because a small amount of mass at a large distance can carry the momentum while the star gradually spins down (*18*). Concurrently the disk plays a role in forming planets, but later than 10 Myear stars seem to lose their protoplanetary material in the little-understood disk transition era. Formation of new planets must cease at this point because the disk mass is insufficient to make a core; if planets have not started to form, they never will (*19*). However, cold rubble piles on the outskirts of the planetary system may continue to slowly bond together, creating Pluto-like bodies around 1000 km in radius (*20*).

The reason for the short lifetimes of protoplanetary disks is uncertain. Protoplanetary disks must be robust because they are common even in very dense regions like the Orion nebula cluster, where stellar encounters must be frequent. Why should disks then suffer a drastic decline in mass after 10 Myear? Stochastic events will operate in star clusters, such as radiative stripping of disks by one very bright star (*21*), but such a star could appear at any time. Perhaps some more general mechanism operates: for example, after 10 Myear a cluster may travel out of the spiral arm where the cloud formed, and so the disk's environment may change as the stars disperse. This discussion reminds us that we mostly observe star formation in action within a few hundred light-years of the Sun and over less than 0.1% of Galactic history, and the processes we see may not be typical. The Milky Way Galaxy has evolved considerably since it formed, processing over 90% of its initial gas into stars and with those stars in turn enhancing the proportions of heavier elements that dominate planetary cores. This means that planets made billions of years ago could have come from rather different disks than those that are presently forming.

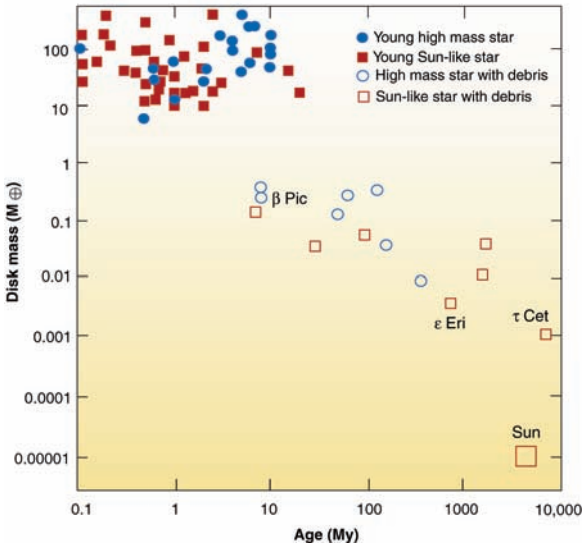


Fig. 2. Disk masses (dust particles only, excluding any gas) in the protoplanetary phase lasting up to 10 Myear and the subsequent debris phase up to 10 Gyear, after (*47*). A brief transition era may link the two phases, in which dust ceases to accumulate but begins to be generated in destructive collisions of planetesimals.

The performance of the disk as a stellar brake is another area of uncertainty. Disk particles will tend to orbit stably or fall in because of drag forces (encountering stellar light and other particles), whereas magnetic fields are the main candidate for outward movement and transport of angular momentum. A field line resists being wound around by the disk's differential rotation (*22*), and so ions tied to the field line may move to a more distant orbit, while neutrals are dragged along by interparticle friction. Other outward forces operate, but, for example, radiation pressure can expel only dust grains in a limited size range. Magnetic fields are in fact deduced to loop between young stars and the inner parts of their disks (*23*) and are inferred in the outer disk from the partial magnetic alignment of elongated grains (*24*). The strengths of some of these fields are poorly known, as is the relative importance of mass transport in the disk and in the stellar wind (*25*). Ultimately, with a wide variety of physical processes operating, the destiny of any one disk particle is unknown, and this affects any predictions for the final mass of the star and the nature of any planets that may form.

The fate of any material not absorbed into the star and planets is also uncertain. Gaps inferred at a few AU in the disks of stars such as GM Aur (*26*) suggest some particles have been absorbed into planets, and more such discoveries are emerging from the Spitzer infrared space telescope (*27*). However, at distances of hundreds of AU, orbital time scales are long and gas densities are low, so this material cannot efficiently form large bodies. Hence, much of the gas component of the disk may disperse rather than being used; less than 10% of the inferred proto-Solar nebula was needed to supply the atmospheres of the gas giant planets. New ground-based observations using a weak line emitted by hydrogen at 2.2 μm (*28*) show that this major component of the gas can persist as late as 8 to 12 Myear (for the star LkCa 15), but here millimeter-wave spectra of trace molecules indicate that the densest gas is ring-like and surrounds a region of lower density (*29*). Photoevaporation by energetic radiation from the young star might contribute to such central clearing (*30*), whereas effects from other stars in a cluster may strip the outer disk (*21*).

Ideas about the structure and environment of the early Sun's disk have strongly influenced models of planet formation. It is now thought that the Sun formed within a star cluster, because of the early presence of isotopes made in supernovae (when massive stars explode) that were incorporated shortly after the Sun formed. Only in a cluster would it be likely that a short-lived massive star would be born around the same time. A cluster environment could explain the rather small scale of the Solar System and its Kuiper Belt (all within about 50 AU), because the disk would be truncated when other stars passed nearby. In the cluster scenario, individual stellar encounters could affect the growth of planets, and so the Solar System may not represent the planetary architecture for all similar stars.

There are two rival models for the formation of the giant planets (Jupiter, Saturn,

Table 1. Characteristics of circumstellar disks. Masses and radii are not exact, because different observational methods trace different disk components and disk structures are not completely spatially resolved. Examples given are stars of Sun-like mass, except in the rare transition disk case exemplified by β Pictoris.

Object phase	Examples and subclasses	Disk lifetime (Myear)	M_{star}:M_{disk}:M_{core} ($M_\odot$)	Disk radius (AU)
Class 0 protostar	VLA 1623 (*49–51*)	0.1	<0.03:0.01:0.5	100 (core is 1000)
T Tauri star (with protoplanetary disk)	HL Tau I GM Aur II V410 Tau III (*19, 26, 52, 53*)	1–10	1:<0.01–0.1:0	100–800
Transition object	β Pictoris (*6*)	10–20	2:10^{-7}:0	500
Main sequence star (with debris disk)	ε Eridani (*48*) τ Ceti (*44*)	100–10,000	1:10^{-9}–10^{-8}:0	50–100

Uranus, and Neptune), whereas terrestrial planets like Mercury, Venus, Earth, and Mars are thought to be made later from leftover planetesimals (*31*). One model for giant planets is growth of a solid core out of dust until this core is massive enough to attract a gaseous atmosphere (*32*). The second model invokes a gravitational instability within the disk (perhaps triggered by a stellar flyby) and spontaneous collapse of part of the disk to make a protoplanet (*33*). The core growth model in its original form (*32*) takes several million years before the atmosphere starts to grow, which may conflict with the observed disk lifetimes (*19*), suggesting mass must be collected more efficiently (*34*). The instability model produces planets very quickly, in only hundreds of years, but requires a substantial disk (more than 10% of the star's mass), which is rarely observed.

About 130 extrasolar planets have now been detected (*35*) by using a technique of very precise spectroscopy that measures the changes in a star's velocity (Doppler wobble) due to the tug on the star by an orbiting planet. The candidate exoplanets are diverse, including hot Jupiters orbiting 10 times closer in than Mercury, bodies in highly eccentric orbits, and masses from 0.05 up to 15 $M_{Jupiter}$ or more. Large masses suggest planets have migrated within the disk in order to find enough raw material, and eccentric orbits may point to encounters between planets, processes that are complex to model. Discriminating between different models requires the detection of a planet in formation, which is very difficult because the feeding zone of a proto-Jupiter in the nearest star formation regions subtends only 0.01 arc sec. Unfortunately, indirect evidence such as gaps in disks is ambiguous, because other processes can let gas flow away even where there is no planet (*36*). We can, however, set some bounds on how many young stars could form analogs to the Sun's planets. The 50 $M_\oplus$ of solids in the terrestrial planets and giant planet cores imply about 0.015 $M_\odot$ of gas existed in a minimum-mass Solar nebula. This is comparatively large, with disks this massive inferred around only 1 star in 10, even in clusters less than a million years old (*37*). This presents a challenge for all the models, because long-term Doppler surveys find 15% or more of stars have some kind of gas giant planet (*38*).

The high frequency of circumstellar disks and increasing rate of planet detections suggest that many stars do develop some kind of planetary system. However, there are some scenarios in which planets cannot form or in which planets are quickly destroyed. Where growth times are long, for example where the disk has little mass or is very spread out, only cometary-sized bodies may form, whereas if growth is very efficient, giant planets may form quickly, migrate rapidly inward, and risk falling into the star. Hence, not all stars may be observed to have planets. The frequency of Earth-mass planets with liquid water is especially uncertain; simulations (*39*) suggest that 5 to 15% among a wide range of roughly Sun-like stars will possess a habitable planet (i.e., with stable temperatures allowing liquid water to exist).

As well as the 15% or more of stars with gas giant exoplanets, another similar-sized population is emerging. These are the debris disk stars, where dust is continually regenerated by the collisions of planetesimals. There is evidence that many of these possess planets, because 80% or more of the disks (*40*) emit only at long wavelengths [far-IR to millimeter], whereas warm grains close to the star would add a mid-IR excess. This lack of a mid-IR excess suggests a central cavity that could be cleared by a giant planet that ejects dust particles passing near its orbit (*41*). If this orbit lies tens of AU from the star, the Doppler wobble is undetectable because it would have a period of hundreds of years, but we can sometimes image the effects of the planets as they shepherd dust particles within the debris belt (Fig. 3).

Our view of debris disks is changing. The phenomenon was found to be rare around stars older than a few hundred Myear (*42*), and so the debris phase might be analogous to Earth's early period with a high-impact rate called the heavy bombardment. Now we find that debris could persist for much of a star's lifetime. A bias from the first far-IR observations was that luminous A-type stars, such as the prototype Vega, heat the dust more and hence have brighter disks. Such stars intrinsically live less than a billion years, but now colder dust emitting at millimeter wavelengths is being detected around longer-lived stars. The debris detection rate does not decline much with age for Sun-like stars (*43*), and in fact the closest Solar analog, τ Ceti, still possesses debris (*44*) after 10 Gyear, more than twice the Sun's age. The debris detection rate of 9% for nearby Sun-like stars (*43*) could be increased to ~30% including the ongoing millimeter searches, and, adding these to the Doppler population of exoplanet candidate systems, as many as 45% of stars could have planetary systems different from the Solar System.

Much new data have been gathered by high-precision spectroscopy in the optical and near-IR and by imaging in the far-IR and millimeter. New facilities are being planned to detect planet formation in action, in particular, the high-resolution millimeter interferometer Atacoma Large Millimetre Array (ALMA) and the sensitive Herschel Space Observatory and James Webb Space Telescope (JWST) that will further explore the IR regime.

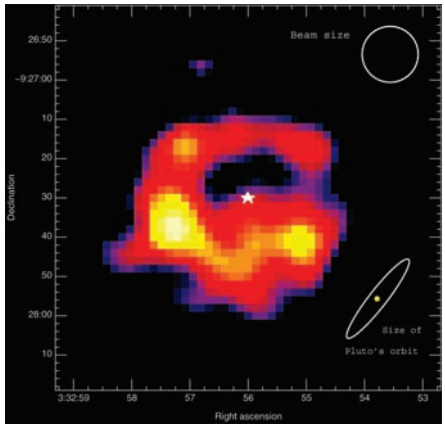


Fig. 3. Image at 850-μm wavelength of thermal emission from dust particles orbiting ε Eridani (*48*); image courtesy of JCMT. The ellipse shows the size Pluto's orbit would appear if it were seen at the star's distance of 10 light-years. The bright spots contain excess dust thought to be held in resonant positions with the orbit of an unseen planet. The positions of the dust peaks can be modeled to reveal the location of the unseen planet (*41*).

References and Notes

1. R. Narayan, E. Quataert, *Science* **307**, 77 (2005).
2. S.-S. Huang, *Astron. J.* **72**, 804 (1967).
3. M. Breger, H. M. Dyck, *Astrophys. J.* **175**, 127 (1972).
4. D. van Blerkom, L. Auer, *Astrophys. J.* **204**, 775 (1976).
5. C. J. Lada, in *Submillimetre Wave Astronomy*, J. E. Beckman, J. P. Phillips, Eds. (Cambridge Univ. Press, Cambridge, 1982), p. 175.
6. B. A. Smith, R. J. Terrile, *Science* **226**, 1421 (1984).
7. I. A. Bonnell, M. R. Bate, H. Zinnecker, *Mon. Not. R. Astron. Soc.* **298**, 93 (1998).
8. K. E. Haisch, E. A. Lada, C. J. Lada, *Astrophys. J.* **553**, L153 (2001).
9. In Keplerian orbit the particle remains at a constant distance r from a point source of gravity with mass M; the circular orbit speed is $(GM/r)^{1/2}$.
10. M. Simon, A. Dutrey, S. Guilloteau, *Astrophys. J.* **545**, 1034 (2000).
11. A. Duquennoy, M. Mayor, *Astron. Astrophys.* **248**, 485 (1991).
12. B. Gladman, *Science* **307**, 71 (2005).
13. M. Landgraf, J.-C. Liou, H. A. Zook, E. Grün, *Astron. J.* **123**, 2857 (2002).
14. The term "dust" can be confusing. It is not meant to imply necessarily organic components, like household dust, or even similar size: Circumstellar dust can be as fine as smoke or as coarse as pebbles.
15. I. M. Coulson, W. R. F. Dent, J. S. Greaves, *Mon. Not. R. Astron. Soc.* **348**, L39 (2004).
16. S. Bontemps, P. André, S. Terebey, S. Cabrit, *Astron. Astrophys.* **311**, 858 (1996).
17. P. Bodenheimer, *Annu. Rev. Astron. Astrophys.* **33**, 199 (1995).
18. L. M. Rebull, S. C. Wolff, S. E. Strom, *Astron. J.* **127**, 1029 (2004).
19. W.-F. Thi *et al.*, *Astrophys. J.* **561**, 1074 (2001).
20. S. J. Kenyon, B. C. Bromley, *Astron. J.* **127**, 513 (2004).
21. D. Johnstone, D. Hollenbach, J. Bally, *Astrophys. J.* **499**, 758 (1998).
22. From (9), there is a difference in velocity between adjacent orbits. The disk thus has differential rotation rather than turning as a solid body.
23. J. M. Oliveira, B. H. Foing, J. Th. van Loon, Y. C. Unruh, *Astron. Astrophys.* **362**, 615 (2000).
24. M. Tamura *et al.*, *Astrophys. J.* **525**, 832 (1999).
25. B. E. Wood, H.-R. Müller, G. P. Zank, J. L. Linsky, *Astrophys. J.* **574**, 412 (2002).
26. W. K. M. Rice, K. Wood, P. J. Armitage, B. A. Whitney, J. E. Bjorkman, *Mon. Not. R. Astron. Soc.* **342**, 79 (2003).
27. NASA press release is available at www.spitzer.caltech.edu/Media/releases/ssc2004-08/release.shtml.
28. J. S. Bary, D. A. Weintraub, J. H. Kastner, *Astrophys. J.* **586**, 1136 (2003).
29. J. S. Greaves, *Mon. Not. R. Astron. Soc.* **351**, L99 (2004).
30. R. D. Alexander, C. J. Clarke, J. E. Pringle, *Mon. Not. R. Astron. Soc.* **354**, 71 (2004).
31. J. E. Chambers, *Earth Planet. Sci. Lett.* **223**, 241 (2004).
32. J. B. Pollack *et al.*, *Icarus* **124**, 62 (1996).
33. A. P. Boss, *Astrophys. J.* **610**, 456 (2004).
34. W. K. M. Rice, P. J. Armitage, *Astrophys. J.* **598**, L55 (2003). In their model, a core random-walks through the disk and so accretes more gas.
35. Online catalogs of planet detections are available at www.obspm.fr/encycl/encycl.html and http://exoplanets.org/.
36. T. Takeuchi, D. N. C. Lin, *Astrophys. J.* **581**, 1344 (2002).
37. D. Nürnberger, R. Chini, H. Zinnecker, *Astron. Astrophys.* **324**, 1036 (1997).
38. D. A. Fischer, R. P. Butler, G. W. Marcy, S. S. Vogt, G. W. Henry, *Astrophys. J.* **590**, 1081 (2003).
39. G. W. Wetherill, *Icarus* **119**, 219 (1996).
40. J. S. Greaves, W. S. Holland, R. Jayawardhana, M. C. Wyatt, W. R. F. Dent, *Mon. Not. R. Astron. Soc.* **348**, 1097 (2004).
41. J.-C. Liou, H. A. Zook, *Astron. J.* **118**, 580 (1999).
42. H. J. Habing *et al.*, *Astron. Astrophys.* **365**, 545 (2001).
43. J. S. Greaves, M. C. Wyatt, *Mon. Not. R. Astron. Soc.* **345**, 1212 (2003).
44. J. S. Greaves, M. C. Wyatt, W. S. Holland, W. R. F. Dent, *Mon. Not. R. Astron. Soc.* **351**, L54 (2004).
45. Image from Hubble Space Telescope was obtained by M. J. McCaughrean, C. R. O'Dell; a related press release is available at http://hubblesite.org/newscenter/newsdesk/archive/releases/1995/45/text/.
46. Image from Gemini Observatory was obtained by R. Jayawardhana and K. Luhman; the press release is available at www.gemini.edu/index.php?option=content&task=view&id=53.
47. M. C. Wyatt, W. R. F. Dent, J. S. Greaves, *Mon. Not. R. Astron. Soc.* **342**, 876 (2003).
48. J. S. Greaves *et al.*, *Astrophys. J.* **506**, L133 (1998).
49. P. André, D. Ward-Thompson, M. Barsony, *Astrophys. J.* **406**, 122 (1993).
50. R. E. Pudritz *et al.*, *Astrophys. J.* **470**, L123 (1996).
51. D. W. Brown *et al.*, *Mon. Not. R. Astron. Soc.* **319**, 154 (2000).
52. D. L. Padgett *et al.*, *Astron. J.* **117**, 1490 (1999).
53. L. A. Hillenbrand, S. E. Strom, F. J. Vrba, J. Keene, *Astrophys. J.* **397**, 613 (1992).

54. J.S.G. is a Norman Lockyer Fellow of the Royal Astronomical Society.

10.1126/science.1101979

REVIEW

The Kuiper Belt and the Solar System's Comet Disk

Brett Gladman

Our planetary system is embedded in a small-body disk of asteroids and comets, vestigial remnants of the original planetesimal population that formed the planets. Once formed, those planets dispersed most of the remaining small bodies. Outside of Neptune, this process has left our Kuiper belt and built the Oort cloud, as well as emplacing comets into several other identifiable structures. The orbits in these structures indicate that our outer solar system's comet disk was shaped by a variety of different physical processes, which teach us about how the giant planets formed. Recent work has shown that the scattered disk is the most likely source of short-period comets. Moreover, a growing body of evidence indicates that the sculpting of the Kuiper belt region may have involved large-scale planetary migration, the presence of other rogue planetary objects in the disk, and/or the close passage of other stars in the Sun's birth cluster.

The appearance of a circumstellar accretion disk is now regarded as a natural part of star formation (*1*). Planets and smaller bodies accrete out of the gas and dusty solids that have been concentrated in this disk close to each forming protostar. Because the process is not 100% efficient, many of the small bodies remain after the completion of the planet-building period; their spatial distribution around our Sun holds clues about the processes that occurred as the planets accreted. In our solar system, the Kuiper belt and other cometary structures give us our best and perhaps only accessible clues about what was going on in our own outer solar system during the final assembly stages of the giant planets.

A central issue in planet formation is confusion about what fundamental process created our giant planets. For the rocky terrestrial planets, the sequence is reasonably well understood. First, the so-called planetesimals, of multikilometer size, were assembled out of whatever solid materials could locally condense in the temperature and pressure conditions at a given solar distance in the nebula. These planetesimals accreted together to form >1000-km bodies, referred to as protoplanetary embryos, which were subsequently put together to form the present planets. Interior to Jupiter's orbit, the process left behind little debris; only the tenuous asteroid belt of unused planetesimals with only ~0.1 to 1% of the region's original mass remains. The planetesimal model for terrestrial planet formation has no competitor because of the wealth of information that constrains it.

In contrast, two very different formation pathways for our giant planets have been discussed. Although it is attractive to theorize that giants form out of large-scale local gravitational instabilities in the disklike solar nebula (*2*), this hypothesis requires a massive and finely tuned accretion disk, with a potentially unrealistically long wait (to allow 1000-km objects like Pluto and the largest trans-neptunian objects to be built)

Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6K 2S5, Canada. E-mail: gladman@astro.ubc.ca

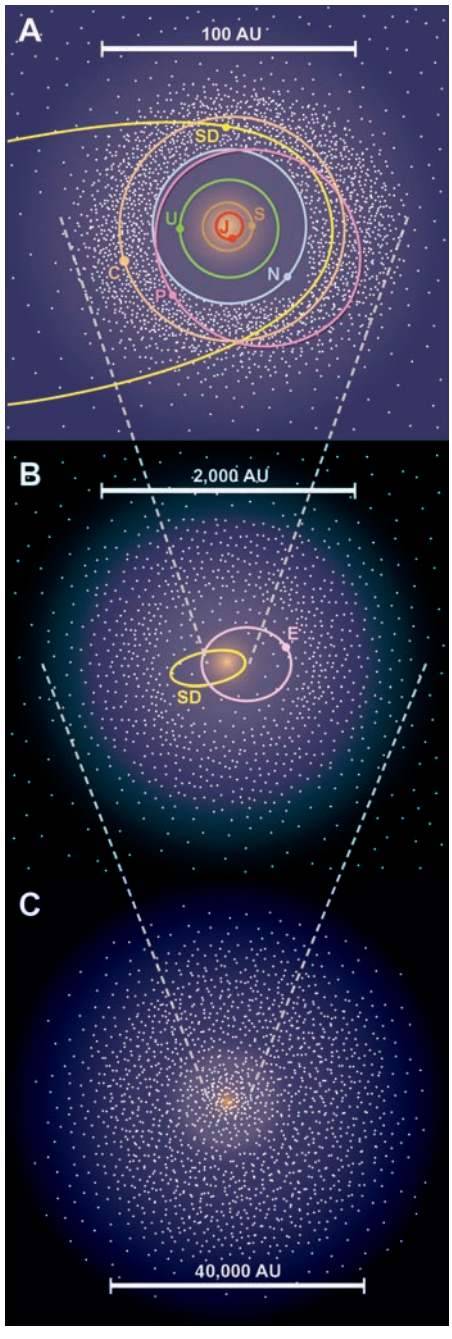


Fig. 1. Schematic of the populations of the solar system's comet disk, looking down from above the plane of the system. (**A**) The main belt just outside the orbits of the four giant planets is the highest spatial density region, in which the classical (C) trans-neptunian objects exist on nearly circular orbits, as well as the plutino objects, which have orbitals sizes similar to that of Pluto (P). The scattered disk (SD) objects have their perihelia in this region as well, but may have aphelia hundreds of astronomical units away, even though they are all detected within about 50 AU of the Sun. J, Jupiter; S, Saturn; U, Uranus; N, Neptune. (**B**) The extended scattered disk (E) objects (some of which have perihelia within 50 AU of the Sun) merge into the inner Oort cloud at semimajor axes around 1000 to 2000 AU. The inner Oort cloud is likely flattened around the solar system's plane (see Fig. 2 for a view perpendicular to this one). (**C**) The roughly spherically symmetric outer Oort cloud, from which Oort cloud comets are dislodged and plunge from all directions into the inner system to become visible active comets.

before the instability occurs. The competitor model posits that the giant planets resulted from the creation of solid icy cores of roughly 10 times the mass of Earth, which then acquired either several Earth masses of nebular gas (in the case of Uranus and Neptune) or about 60 to 300 Earth masses of gas (in the case of Saturn and Jupiter) (*3*); this core-accretion model requires the correct timing of certain events and has problems during the final stage of giant planet assembly (*4*). Studies of extrasolar planetary systems have yielded new insights, but interpretation is complicated because the process yielding the generally massive extrasolar planets may be different from that which yielded our own. Gravitational instability followed by migration works well in producing so-called "hot Jupiters," but our solar system may have used core accretion to build at least some of its giant planets.

Constraining giant planet formation is difficult because of the subsequent evolution of the planets and the difficulty of sampling them directly. However, the orbital distribution of the remaining small objects in the outer solar system was established by the gravitational influence of the forming giant planets. Studying this orbital distribution teaches us what processes were forming the planets and sculpting the outer solar system.

Although most people are familiar with active comets in the form of nearby kilometer-scale icy objects with huge tails that are created by volatilization within about 2 astronomical units (AU) of the Sun, comets in this class have only a fleeting existence because they are quickly ejected from the solar system by Jupiter or perhaps lose all their surface ices and transform themselves to a more asteroidal appearance. The active comets and the more distant Centaur population (cometary objects crossing the orbits of the giant planets but not yet approaching the Sun) are members of a transitory population that must be fed by a more distant source. Although the existence of the Kuiper belt and Oort cloud were deduced from the orbital properties of the active comet distribution, we can now directly study the former reservoir observationally, and we will not further discuss the transitional populations.

Outside of Neptune's circular orbit at 30 AU, we find the first cometary reservoir: the Kuiper belt. This portion of the solar system's remaining small-body disk consists of a main belt, a scattered disk, and an extended scattered disk (Fig. 1). These are collectively referred to as trans-neptunian objects, although the boundaries of the various populations are only loosely defined (*5*).

The main belt is the region of nearly circular orbits with semimajor axes a = 35 to 56 AU, with some orbits in mean-motion resonance with Neptune (*6*). This region has been heavily depleted if formed in place, given that it contains only ~0.1 to 1% of the mass needed to accrete the 1000-km objects seen (*7*). The orbits are so sufficiently eccentric and inclined that other processes must have affected the main belt, allowing the mass depletion and pumping the random velocities (*8*). The inclination distribution of the main belt shows a flatter dynamically cold component immersed in a fatter torus of higher inclination objects (*9*), which cannot be convincingly explained by any current model. This two-component structure may be related to a color correlation in which the cold component of the belt appears to have redder optical color (*10*). The cold (low-eccentricity and low-inclination) component seems to end at roughly the position of the 2:1 mean-motion resonance with Neptune (*11*, *12*), which might be explained by the migration of Neptune into the outer solar system (*13*).

The scattered disk consists of objects on highly eccentric orbits of large semimajor axis with perihelia q (the closest approach to the Sun) near Neptune in the range of 30 to 38 AU (*14*). Because at their aphelia (the farthest point from the Sun) they may be 100 to 3000 AU away, these objects are only visible for a small fraction of their orbital period while near perihelion. They are probably Kuiper belt comets or planetesimals that originally formed between the giant planets, which were flung outward to long-lived but eventually unstable orbits outside of Neptune. They are slowly being eroded by continued interaction with Neptune (*15*), although some recent resupply comes from Kuiper belt objects slowly leaking out of the main belt region. Although now reduced to ~1% of the population 4 billion years ago, the scattered disk remains the most likely reservoir for many of the active comets (*16*).

The extended scattered disk objects also have large semimajor axis orbits and high eccentricities but with perihelion q > 38 AU (*17*) (this lower limit on q depends somewhat on the semimajor axis). Such orbits are highly stable against being eliminated by close encounters with Neptune over the age of the solar system, but they also cannot be produced by such encounters.

Outside of the Kuiper belt we find the Oort cloud, a multicomponent structure containing huge numbers of comets that were ejected from the planetary region to thousands or tens of thousands of astronom-

ical units (initially on what would be termed scattered disk orbits) and then had their perihelia raised. This perihelia raising could be accomplished by the gravitational tidal effect of the galaxy (*18*) or by encounters with the many passing stellar neighbors that the Sun might have had in the cluster in which our star formed (*19*). The latter process is ineffective now because of the low stellar densities in the Sun's galactic environment, but an inverse encounter process allows passing stars or giant molecular clouds to return comets to the planetary system and potentially become active. This process is only effective in the outer Oort cloud, where the speed changes induced by a typical stellar flyby (at distances measured in tens of thousands of astronomical units) are sufficiently large relative to the comet's orbital speed to allow it to have nearly zero angular momentum and be transferred to an orbit that will come within a few astronomical units of the Sun. In contrast, the inner Oort cloud (with semimajor axes from about 2000 to 20,000 AU) is too tightly bound for this process to be efficient; the inner Oort cloud is thus still inaccessible by direct means (because the comets in it are too faint to be seen) and indirect means (because the comets in it cannot be perturbed onto orbits where they will become visible).

Although the Kuiper belt was initially hypothesized as the remnant band of planetesimals outside the orbit of Neptune (similar to the asteroid belt between Mars and Jupiter), we now realize that the scattered and extended scattered disks may both have populations exceeding that of the main belt. The creation and subsequent erosion of the two scattered populations over the age of the solar system has been governed by the movements of much more massive objects moving near or through the distribution.

A strawman model for the formation of the main belt and subsequent creation of the scattered disk hypothesizes that the giant planets were formed essentially where they are now, in the midst of a disk of planetesimals on nearly circular orbits from Jupiter to beyond Neptune (Fig. 2). These planetesimals have their orbits perturbed by the weak gravitational forces from the giant planets and eventually begin having a series of stronger gravitational scatterings once their orbits cross a planetary orbit; these planetesimals are then either flung outward into the scattered disk or, if the planetary interaction is strong enough, ejected from the solar system on hyperbolic orbits, never to return. Some of the scattered disk objects have aphelia out far enough (>2000 AU) that external processes (such as passing stars or the tidal effects of our galaxy) can emplace them into the Oort cloud; a few return in the form of the Oort cloud (isotropic) comets (*20*).

The scattered disk is unstable due to the ability of close encounters with Neptune (over time scales of hundreds of millions of years) to remove objects by increasing their semimajor axes to very large values (either emplacing them in the Oort cloud or ejecting them from the solar system) or by gravitational encounters eventually pulling them down deeply into the planetary region at which point the other planets also begin to alter orbits markedly. These "pulled-down" objects with semimajor axes of less than 30 AU are the Centaurs, and although some are able to return to the scattered disk, most eventually have strong close encounters with Jupiter and Saturn and are ejected after tens of millions of years, with a fraction briefly becoming the active Jupiter-family comets before their demise (*21*). Indeed, it seems likely that the scattered disk is where the majority of observed Jupiter-family comets spent most of their lives in the solar system; the erosion rate from the scattered disk is more efficient than from the main belt for comparable populations (*15*, *16*). Just after their formation, the giant planets would have flung trillions of cometary planetesimals out into the scattered disk. After this initial colonization epoch, the number of scattered disk objects decreases with time, and so in the strawman model, the population would have been the highest early in the solar system and would have monotonically declined afterward. The current population of the scattered disk is about 1% of the population 4.4 billion years ago, under the reasonable assumption that the current input into the scattered disk is small (*15*).

During or after the period in which the giant planets reached their current masses, the main Kuiper belt was eroded (Fig. 2); it lost >99% of its mass, and its orbital distribution was modified to higher orbital eccentricities and inclinations. The time scale for this was likely, at most, a few hundred million years and may have been much shorter. The initial buildup of the scattered disk occurred simultaneously. Populating the outer and inner Oort cloud probably occurred on a longer time scale; the tidal effects operate on a time scale of the age of the solar system (*18*), but the perihelion lifting into the inner Oort cloud during an early passing star phase would have occurred only in the first few hundred million years (*19*).

The strawman model above is known to be lacking in several ways. The mass erosion and orbital excitation of the main belt cannot be simply accomplished by the gravitational perturbations of the four giant planets. In addition, the main belt has a complex structure with a two-component inclination distribution, objects trapped in mean-motion resonances (*22*), and a surface-density dropoff at ~48 AU. The scattered disk's orbital distri-

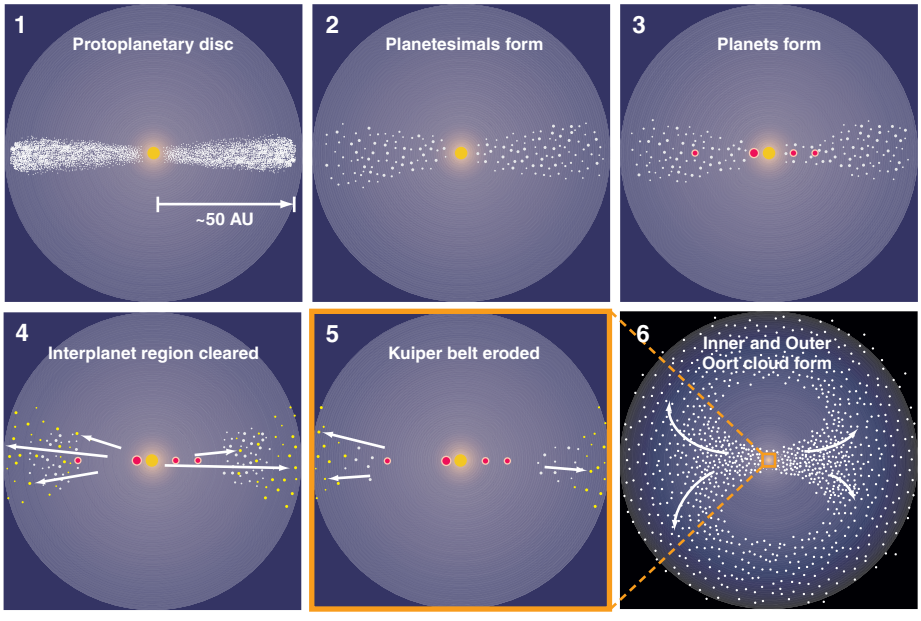


Fig. 2. Strawman model of the evolution of the solar system's comet reservoirs. (1) A flattened accretion disk of gas and dust generates (2) planetesimals through solid-body accretion. The Kuiper belt region is built during this period. The outer planets form (3) in this planetesimal disk and rapidly (in 1 to 10 million years) deplete the interplanet region (4) by hurling the planetesimals out to the scattered disk. On a similar or slightly longer time scale, the Kuiper belt exterior to Neptune is eroded (5) and has its orbital distribution excited. Most of the dispersed objects are placed in the scattered disk (yellow) with large semimajor axes. Those with aphelia thousands of astronomical units away are perturbed by galactic tides and/or passing stars into the inner Oort cloud, whereas those with distances out beyond 10,000 AU are dispersed into the roughly spherical outer Oort cloud (6).

bution seems to match that which should be produced by scattering with the giant planets (*23*), but the extended scattered disk is not populated with enough efficiency (Fig. 3) to create it in the strawman model (*17*).

After a gravitational encounter with a planet, a small body will be on a Keplerian orbit, the perihelion distance of which cannot be farther than the distance at which the orbit-changing encounter occurred. As such, it is very difficult for perihelion distances q to reach values above about 38 to 40 AU (*17*). Thus, the scattered disk extends out to very large semimajor axes and heliocentric distances but is tied to the planetary region; scattered disk objects return every orbit at their closest solar passage to a perihelion $q < 40$ AU. There are a few mean-motion resonances in the scattered disk regime that allow objects (*15*) to temporarily have their q raised (Fig. 3), but in the current solar system such states are always transient.

Thus, the population of objects with perihelia $q > 40$ AU and semimajor axes $a < 2000$ AU does not fit into the strawman model of an outer solar system with a vestigial Kuiper belt in the region of 35 to 50 AU and a scattered disk with large semimajor axis but perihelion in the planetary region. The first solid evidence was provided by an exceptional object with the designation 2000 CR105 (*24*), tracking of which established an orbital semimajor axis of about 220 AU (which itself is not a surprise) but a perihelion distance of 44 AU (*17*). Several other such objects now have been identified (Fig. 3), including most recently the orbit of minor planet Sedna with a semimajor axis of slightly more than 500 AU with a perihelion at a stunning 76 AU away from the Sun (*25*). Because such objects cannot be produced by repeated scattering by the giant planets, the term "extended scattered disk" was coined for the $q > 40$ AU population, whose semimajor axes are small enough that the perihelion cannot have been raised by the effect of galactic tides (*18*). Because extended scattered disk objects are only visible for a tiny fraction of their orbit (their feeble reflected light fades below visibility over 99% of their orbit), one estimates that there must be more than a million such objects larger than 100 km (*17*), which is more populous than either the main belt or scattered disk. Depending on the size distribution, there could be more than an Earth mass (*25*) of material present. An exact population estimate is difficult, because it depends on the unknown orbital distribution (in particular, it is very sensitive to the distribution of perihelion distance, where the objects are most visible).

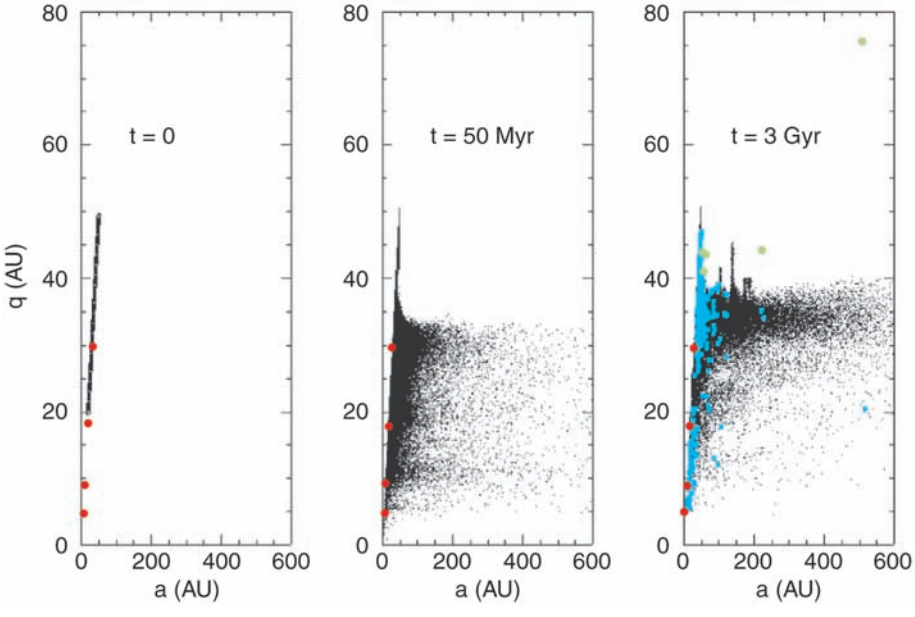


Fig. 3. The production of the scattered disk by the giant planets. In this calculation, a disk of planetesimals is placed on initially circular orbits in the midplane of the outer solar system, from 20 to 50 AU [diagonal line of black points, $t = 0$ panel (**left**)]. The planets (red points) have their current orbits and masses. After only 50 million years (**middle**), the planets have flung many small bodies out to semimajor axes of several hundred astronomical units and transferred others down to Centaur orbits near Jupiter and Saturn ($a < \sim 30$ AU) where their lifetimes against ejection from the solar system are $<\sim 10$ million years. By 3 billion years (**right**), much of the interplanetary region ($q < 30$ AU) is being cleared out and the scattered disk ($q = 30$ to 38 AU) has been built. Production by means of scattering (black points) of objects with large a and $q > 38$ AU (in the extended scattered disk) are rare; two examples are visible, both due to temporary trapping in mean-motion resonances with Neptune. Green points show five known extended scattered disk objects, and the blue points show other known Kuiper belt objects in the main belt, scattered disk, and Centaur populations.

Three scenarios have been proposed to form the extended scattered disk. The first is the possibility of slow diffusive chaos near the borders of the mean-motion resonances, allowing particles that get temporarily trapped in the resonances to exist for long periods at high q (*15*). It has been shown (*17*) that resonant perihelion lifting could not explain the orbit of objects such as 2000 CR105 in sufficient numbers to make the production of the well-populated extended scattered disk feasible. Thus, cosmogonic models (related to the formation period of our solar system) are needed.

The second scenario uses a stellar passage to modify the orbits of scattered disk objects. Although close (<1000 AU) stellar encounters are unlikely to have happened in the last few billion years, they are plausible if the Sun formed, as commonly occurs, in a dense stellar cluster (*26*). If the solar system had time to first build a planetesimal population and then accrete giant planets to fling many planetesimals out into a scattered disk, then a subsequent near-stellar passage (at about 200 AU) seems sufficient to lift the perihelia of many objects (*27*, *28*). Alternatively, one might try to build a large-diameter planetesimal population directly at 50 to 200 AU (*29*), which are then scattered to elliptical orbits. The tricky part of these hypotheses is the timing: The stellar passage must occur late enough that huge numbers of planetesimals have formed and are in the right place to have their perihelia raised, and yet the passage must not occur so late as to destroy the scattered disk and the Oort cloud, which then lack a resupply population in the form of remaining comets closer to the Sun (*30*). In addition, these models (*27*, *28*) have confined themselves to an advantageous encounter geometry with low encounter angle with the planetesimal disk, whereas a relative inclination of 90° is the most probable. The probability that encounters at random inclinations and impact parameters produce a final configuration similar to our scattered and extended scattered disks needs to be established.

The third scenario invokes the presence of one or more rogue planets the size of Mars, Earth, or even larger in the early outer solar system. In core-accretion models, the roughly 10 to 15 Earth-mass cores of the giant planets are built up from planetesimals (*31*), but the planet-formation process was not perfectly efficient. The building blocks of these cores may be Earth-mass objects near the "oligarchy mass" of solid bodies that would have been built in an initial burst of planetesimal accretion in the outer solar system, but mass growth of which has not succeeded as well as the giant planet cores. These failed planets will be scattered around the outer solar system by the now-dominant giant planets and can modify

the orbital distribution of the initially quiescent Kuiper belt distribution. Early models attempted to excite the main belt's random orbital velocities this way and remove most of its objects (*32*), and other studies have tried to explain various aspects of the Kuiper belt in such a context (*33*, *34*). The rogue planet is usually eventually scattered out of the solar system, although there exists the possibility (*17*) that one or more could still survive and remain undetected in the distant portions of the scattered or extended scattered disks. A recent study (*28*) concluded that rogue planet scenarios produce an observational signature that is in conflict with the observational distribution of the extended scattered disk and that a stellar passage model is more likely. Other work shows that the secular interaction of the rogue planet with the scattered disk allows efficient perihelion raising (*35*). Both scenarios warrant further work.

The study of the minor planets in the Kuiper belt has yielded constant surprises both observationally and theoretically over the past 5 years. In few other fields in astronomy is there such a tight and rapid advance in both the developments coming from telescopic work (driven by improving detector technology) and computational dynamics (enabled by evolving computer hardware). Ten years from now, our perspective on the solar system's small-body disk will be considerably advanced by a more complete census of our outer solar system and the coming ability to directly probe the region of Kuiper belts around other stars promised by the next generation of telescopes.

References and Notes

1. J. S. Greaves, *Science* **307**, 68 (2005).
2. A. Boss, *Astrophys. J.* **536**, L101 (2000).
3. G. Wuchterl, T. Guillot, J. Lissauer, in *Protostars and Planets IV*, V. Mannings, A. Boss, S. Russell, Eds. (Univ. Arizona Press, Tucson, AZ, 2000), pp. 1081–1109.
4. B. Gladman, *Nature* **396**, 513 (1998).
5. B. Gladman, *Highlights Astron.* **12**, 193 (2002).
6. The semimajor axis is the average orbital distance (half the long axis of the elliptical orbit). Mean-motion resonances occur when the orbital period of the trans-neptunian object is near a fractional multiple of a planet. For example, objects near $a = 39.4$ AU may be in the 3:2 mean-motion resonance with Neptune; Pluto and the so-called plutino population are examples. The 2:1 resonance with Neptune at $a = 48$ currently marks the outer edge of known objects with low orbital eccentricity.
7. S. Stern, J. Colwell, *Astrophys. J.* **490**, 879 (1997).
8. B. Gladman, J. Kavelaars, J.-M. Petit, A. Morbidelli, M. Holman, T. Loredo, *Astron. J.* **122**, 1051 (2001).
9. M. E. Brown, *Astron. J.* **121**, 2804 (2001).
10. A. Doressoundiram, *Earth Moon Planets* **92**, 131 (2003).
11. R. L. Allen, G. Bernstein, R. Malhotra, *Astrophys. J.* **549**, L241 (2001).
12. C. Trujillo, M. E. Brown, *Astrophys. J.* **554**, L95 (2001).
13. R. Gomes, A. Morbidelli, H. Levison, *Icarus* **170**, 492 (2004).
14. C. Trujillo, D. Jewitt, J. Luu, *Astrophys. J.* **529**, L103 (2000).
15. M. J. Duncan, H. F. Levison, *Science* **276**, 1670 (1997).
16. G. Bernstein *et al.*, *Astron. J.* **128**, 1364 (2004).
17. B. Gladman *et al.*, *Icarus* **157**, 269 (2002).
18. M. Duncan, T. Quinn, S. Tremaine, *Astron. J.* **94**, 1330 (1987).
19. J. A. Fernandez, A. Brunini, *Icarus* **145**, 580 (2000).
20. P. Wiegert, S. Tremaine, *Icarus* **137**, 84 (1999).
21. H. Levison, M. Duncan, *Icarus* **127**, 13 (1997).
22. E. Chiang, A. Jordan, *Astron. J.* **124**, 3430 (2002).
23. A. Morbidelli, V. Emel'yanenko, H. Levison, *Mon. Not. R. Astron. Soc.* **355**, 935 (2004).
24. M. Buie *et al.*, *Earth Moon Planets* **92**, 113 (2003).
25. M. Brown, C. Trujillo, D. Rabinowitz, *Astrophys. J.* **617**, 645 (2004).
26. F. Adams, G. Laughlin, *Icarus* **150**, 151 (2001).
27. S. Ida, J. Larwood, A. Burkert, *Astrophys. J.* **528**, 351 (2000).
28. A. Morbidelli, H. Levison, *Astron. J.* **128**, 2564 (2004).
29. S. J. Kenyon, *Publ. Astron. Soc. Pac.* **114**, 265 (2002).
30. H. Levison, A. Morbidelli, L. Dones, *Astron. J.* **128**, 2553 (2004).
31. P. Goldreich, Y. Lithwick, R. Sari, *Annu. Rev. Astron. Astrophys.* **42**, 549 (2004).
32. J.-M. Petit, A. Morbidelli, G. Valsecchi, *Icarus* **141**, 367 (1999).
33. J. A. Fernandez, W. H. Ip, *Icarus* **58**, 109 (1984).
34. A. Brunini, M. Melita, *Icarus* **160**, 32 (2002).
35. B. Gladman, E. Nadal, C. Chan, *Bull. Am. Astron. Soc.* **36**, 08.05 (2004).

10.1126/science.1100553

VIEWPOINT

From Stars to Dust: Looking into a Circumstellar Disk Through Chondritic Meteorites

Harold C. Connolly Jr.

One of the most fundamental questions in planetary science is, How did the solar system form? In this special issue, astronomical observations and theories constraining circumstellar disks, their lifetimes, and the formation of planetary to subplanetary objects are reviewed. At present, it is difficult to observe what is happening within disks and to determine if another disk environment is comparable to the early solar system disk environment (called the protoplanetary disk). Fortunately, we have chondritic meteorites, which provide a record of the processes that operated and materials present within the protoplanetary disk.

Chondrites are 4.5672 ± 0.6 billion-year-old (*1*) rocks derived from the aggregation of dust and other rocks within the protoplanetary disk (Fig. 1). Arguably the oldest components of chondrites are calcium- and aluminum-rich inclusions [CAIs (*2*)] that contain mineral phases predicted to be the first to condense from a gas of solar composition. These objects range in size from submillimeter to centimeter, the largest being igneous rocks (melted and crystallized). The dominant component of chondrites is chondrules—submillimeter- to millimeter-sized igneous rocks that are probably younger than CAIs. Chondrules have compositions less refractory than those of CAIs and are mostly composed of the Mg- and Fe-rich silicate minerals, olivine and pyroxene, with varying concentrations of glass (*3*, *4*).

From an astrophysical viewpoint, if the igneous components of chondrites did not exist, they would not be predicted to exist. They are argued to have formed as free-floating objects within the earliest stages of the protoplanetary disk before planet formation—melted and quenched by some unknown mechanism or mechanisms (*4*). It is important to understand how, when, and why they formed, because this will lead to an understanding of the early conditions in the protoplanetary disk and how planets grew and evolved from these chondritic building blocks.

The Chondrites and the Protoplanetary Disk workshop in Hawaii (*5*) brought together about 150 participants from meteoritics, cosmochemistry, planetary science, astronomy, and astrophysics to understand how chondrites formed and how their formation affected the evolution of the solar system. One key question addressed at this workshop was whether the Sun formed in an isolated environment or in a cluster, where contamination from other stellar systems would be possible. As Greaves discusses in her Review article (*6*), evidence for the existence of a radioactive isotope of aluminum, ^{26}Al, within the protoplanetary disk suggests that the Sun likely formed within a cluster of stars. The evidence of live ^{26}Al and other short-lived radionuclei such as ^{60}Fe (*7*) comes from

Department of Physical Sciences, Kingsborough College and the Graduate School of the City University of New York, Brooklyn, NY 11235, USA; Department of Earth and Planetary Sciences, American Museum of Natural History, New York, NY 10024, USA; and Department of Geological Sciences, Rutgers University, Piscataway, NJ 08854–8066, USA. E-mail: hconnolly@kbcc.cuny.edu

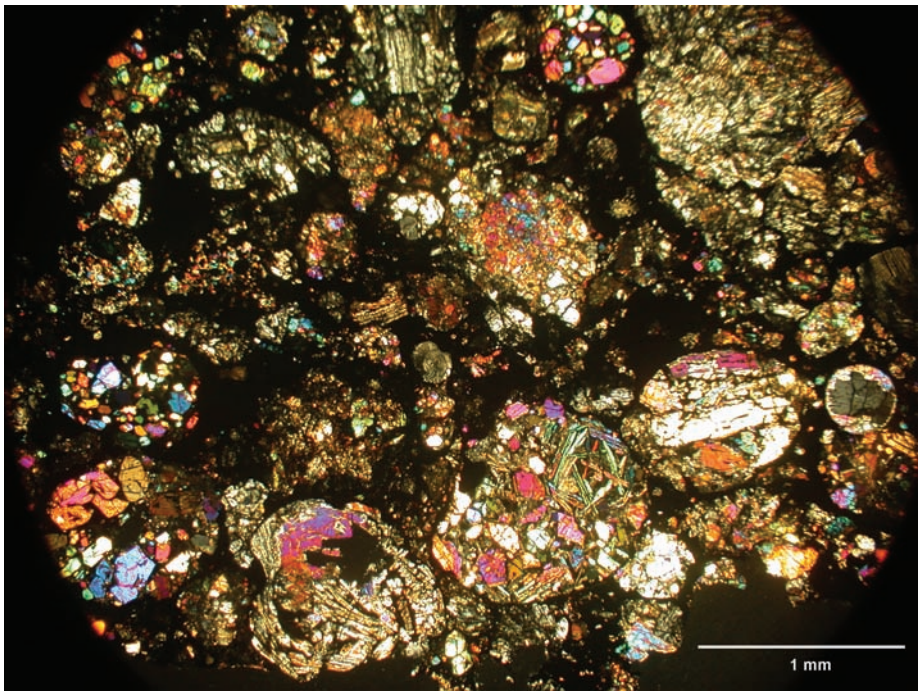


Fig. 1. Crossed polarized light image of a thin section of the ordinary chondrite Chainpur. The round objects are the chondrules, filled with angular minerals (olivine and pyroxene) and glass, whereas the finer-grained and darker materials are the matrix. CAIs are generally found within the matrix but are very rare in ordinary chondrites and too small to be seen in this image.

chondrites. The observed abundance of ^{60}Fe and its inferred initial abundance require a supernova source in the vicinity of the Sun at or slightly after the contraction phase of the solar molecular cloud (solar nebula). Some short-lived radionuclei, such as ^{10}Be, may have been produced through spallation reactions in the solar system (*8*), although these radionuclei may also be explained as cosmic rays trapped within the solar nebula (*9*). Models (*8*) have been proposed to extend the production of most short-lived isotopes through irradiation from energetic solar particles within the solar system, but such results are not widely favored by meteoriticists. Thus, some radioactive nuclei, particularly ^{26}Al and ^{60}Fe, require a nearby supernova explosion to contaminate the solar nebula, whereas other nuclei may be produced within the evolving solar system.

Formation of the solar system within a star cluster may have affected the disk's evolution in other ways. It has been suggested (*10*) that close encounters with a rogue star during the earliest stages of formation may have truncated the disk outside of Neptune's orbit. Gladman in his Review article (*11*) suggests that such a process may have had an impact on the formation of Kuiper Belt objects. There is little evidence from chondrites to address this idea. One exception may be CAIs known as FUN inclusions, which contain fractionated unidentified nuclear isotopic anomalies (*2*). Although speculation, they or their precursors might have been injected into the disk during a close pass with a rogue star, and thus FUN inclusions could predate other CAIs.

Non–mass-dependent variations in the stable isotopes of oxygen (^{16}O, ^{17}O, and ^{18}O) are observed within CAIs and chondrules. CAIs are generally enriched in ^{16}O compared with chondrules (*12*). Until recently, the major cause of this difference was thought to be a gas-solid exchange process in the solar nebula. In this model, light oxygen was inherited within solids from a stellar source—injected into the solar nebula or young disk—and during high-temperature processing of solids, the oxygen isotopes were exchanged between the gas and the modified solids. However, the hypothesis now in favor attributes the oxygen isotopic variations to effects of photodissociation and photo-evaporation (*12*). The source of the evaporation could have been the young Sun or another stellar source close to the solar nebula and/or young disk (*12*, *13*).

CAIs and chondrules provide two other important clues to the protoplanetary disk. First, the environment in which CAIs formed was different from that of chondrules. CAIs have different oxygen and magnesium isotopic compositions (the latter suggesting mass-dependent fractionation due to evaporation), and petrographic and chronologic data indicate that they melted at different times (*7*, *14*). CAIs likely formed within a single environment poor in dust with few proto-CAIs that were later scattered within the disk, whereas chondrules likely formed in areas rich in dust and proto-chondrules. This conclusion does not, however, require different melting mechanisms for the objects. It does mean that the dust that eventually formed the building blocks of planets varied within the disk. Second, chondrule formation lasted for 2 to 5 million years (*7*, *14*). The protoplanetary disk must have lasted at least this long. Here the rock record of processes within a disk agrees with observations of the lifetimes of protoplanetary disks (*15*). Thus, the meteorite record and observations of the cosmos are converging on the lifetime of disks and suggest solar system formation within a stellar cluster.

References and Notes

1. Y. Amelin *et al.*, *Science* **297**, 1678 (2002).
2. G. J. MacPherson *et al.*, in *Meteorites and the Early Solar System* (Univ. of Arizona Press, Tucson, AZ, 1988), p. 746.
3. A. J. Brearley, R. H. Jones, *Rev. Mineral.* **36**, 3.1 (1988).
4. H. C. Connolly Jr., S. J. Desch, *Chem. Erde* **95**, 95 (2004).
5. A. Krot *et al.*, in *Chondrites and the Protoplanetary Disk* (School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, HI, 2004), p. 232.
6. J. S. Greaves, *Science* **307**, 68 (2005).
7. N. T. Kita *et al.*, in *Chondrites and the Protoplanetary Disk* (School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, HI, 2004), p. 97.
8. M. Gounelle *et al.*, *Astrophys. J.* **548**, 1051 (2001).
9. S. J. Desch *et al.*, *Astrophys. J.* **602**, 528 (2004).
10. B. Reipurth, in *Chondrites and the Protoplanetary Disk* (School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, HI, 2004), p. 173.
11. B. Gladman, *Science* **307**, 71 (2005).
12. K. D. McKeegan *et al.*, in *Chondrites and the Protoplanetary Disk* (School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, HI, 2004), p. 123.
13. J. J. Hester *et al.*, *Science* **304**, 1116 (2004).
14. A. M. Davis *et al.*, in *Chondrites and the Protoplanetary Disk* (School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, HI, 2004), p. 27.
15. K. E. Haisch *et al.*, *Astrophys. J. Lett.* **553**, L153 (2001).

10.1126/science.1108284

Black Hole Accretion

Ramesh Narayan[1]* and Eliot Quataert[2]*

Black holes are most often detected by the radiation produced when they gravitationally pull in surrounding gas, in a process called accretion. The efficiency with which the hot gas radiates its thermal energy strongly influences the geometry and dynamics of the accretion flow. Both radiatively efficient thin disks and radiatively inefficient thick disks are observed. When the accreting gas gets close to the central black hole, the radiation it produces becomes sensitive to the spin of the hole and the presence of an event horizon. Analysis of the luminosities and spectra of accreting black holes has yielded tantalizing evidence for both rotating holes and event horizons. Numerical simulations imply that the relativistic jets often seen from accreting black holes may be powered in part by the spin of the hole.

Accretion is the term used by astrophysicists to describe the inflow of matter toward a central gravitating object or toward the center of mass of an extended system. Accretion is one of the most ubiquitous processes in astrophysics and is responsible for forming much of the structure we see around us. Galaxies formed early in the universe as gas flowed in toward the center of gravitational potential wells established by dark matter (*1*). Stars form inside galaxies even today as gas clouds collapse and fragment under their own self-gravity and then grow by accreting surrounding gas (*2*). Planets—including Earth—form as gas and rocks coalesce in the debris surrounding a new star (*3*). Perhaps the most spectacular observational manifestations of accretion occur, however, when the central object is a black hole.

Black holes are among the most striking predictions of Einstein's theory of General Relativity: So much mass is compressed into such a small volume that gravity overwhelms all other forces, and nothing—not even light—can escape. Instead of a normal surface, a black hole has an event horizon, a virtual surface that separates the outside world from the region of the black hole from which nothing escapes. Remarkably, General Relativity predicts that macroscopic black holes such as those that are studied in astrophysics are extremely simple objects; they can be completely described by just two parameters: their mass and spin (rotation rate) (*4*–*6*).

Astronomers now recognize that black holes come in many varieties, from holes with masses a few times $M_\odot$, the mass of the Sun ($\sim 2 \times 10^{33}$ g), to supermassive holes with masses between 10^6 and 10^9 $M_\odot$. Whether there are intermediate-mass black holes between these two extremes remains uncertain and is the focus of much current research (*7*). Black holes with masses of ~ 10 $M_\odot$ form when stars in the range ~ 30 to 100 $M_\odot$ exhaust their nuclear fuel and collapse under their own weight (*8*); there are perhaps 10 million such black holes in a galaxy like the Milky Way. Most are dark and invisible, but some become powerful sources of x-rays when they accrete matter from a companion star; these are called x-ray binaries (XRBs). More speculatively, the accretion of matter onto a newly formed black hole at the center of a collapsing star may give rise to gamma-ray bursts, the most energetic explosions in the universe (*9*).

In addition to these smaller holes, there is now compelling evidence that nearly every galaxy contains a much more massive black hole at its center. These range in mass from $\sim 10^6$ to 10^9 $M_\odot$, with the mass of the central black hole being closely tied to that of the bulge (*10*) of its host galaxy, $M_{\rm BH} \approx 10^{-3}\ M_{\rm bulge}$ (*11*–*13*). The most direct evidence for such a massive black hole is at the center of the Milky Way, where the elliptical orbits of stars around the black hole have been measured, allowing a precise determination of its mass $\approx 3.7 \times 10^6\ M_\odot$ (*14*, *15*). When one of these massive black holes accretes gas from its surroundings, it is called an active galactic nucleus (AGN). An AGN can be extremely luminous (up to 10^{48} erg s^{-1}), outshining all of the stars in its host galaxy.

Although the existence of central massive black holes is empirically well established, their formation process remains poorly understood. The tight connection between black hole mass and galaxy mass provides a strong hint, however, and suggests that the formation of a massive black hole is intimately tied to the formation of the host galaxy in which it resides. A promising explanation for this coupling is that as the black hole grows by accretion, it deposits sufficient momentum and energy into its surroundings to blow gas out of the galaxy, thereby shutting off its own fuel supply (*16*–*18*). This process may also determine the total number of stars formed in a given galaxy and may regulate structure on scales as large as those of clusters of galaxies (*19*, *20*). Thus, far from being mere curiosities, black holes may actually play an active role in the growth and evolution of structure in the universe.

Accretion Physics

Gas flowing in toward a central object almost inevitably has angular momentum that prevents it from directly reaching the object. Instead, the gas settles into a disk-like structure whose orientation is defined by the angular momentum of the gas (Fig. 1). This simple physics is responsible for the prevalence of disks in astrophysics, in systems as diverse as the disk of gas that defines the Milky Way and the disk of planets that defines the solar system. Given the barrier posed by angular momentum, some form of friction is required to allow gas to flow in toward the central object. Magnetic forces in an ionized plasma probably supply this friction in many systems, efficiently transporting angular momentum outward and allowing accretion to proceed (*21*, *22*).

In addition to transporting angular momentum, magnetic stresses also convert some of the gravitational potential energy of the accreting gas into heat. The processes by which this occurs are complex and poorly understood, but the amount of energy released does not depend sensitively on these details and can be reliably estimated. The dynamics of the resulting accretion disk then depends critically on whether this thermal energy is radiated away. We can divide accretion flows into two major classes on the basis of their radiative efficiency $\varepsilon \equiv L/\dot{M}c^2$, where L is the luminosity produced by the accretion flow, $\dot{M}$ is the mass accretion rate, and c is the speed of light (and thus $\dot{M}c^2$ is the rate at which rest mass energy is accreted).

If the dissipated energy is radiated away on a time scale shorter than the time it takes the gas to flow into the black hole, the gas cools rapidly and settles into a thin accretion disk (the vertical thickness of the disk is typically in the range 0.1 to $\sim$3% of the radius, i.e., much thinner than the simulation shown in Fig. 1). The efficiency ε is then $\sim$0.06 to 0.4, depending on the spin of the black hole; thus, thin accretion disks around black holes are the most efficient known source of power in the universe, up to $\sim$50 times as efficient as nuclear fusion in stars (*23*). By contrast, if

[1]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA. [2]Astronomy Department, University of California, 601 Campbell Hall, Berkeley, CA 94720, USA.

*To whom correspondence should be addressed. E-mail: narayan@cfa.harvard.edu, eliot@astron.berkeley.edu

the gas cannot radiate its energy away, it forms a hot thick accretion disk (*24*) with $\varepsilon \ll 1$ (the disk thickness is then 20% or more of the radius, as in Fig. 1).

Accretion onto a black hole can take either of these two forms (Fig. 2). At low accretion rates $\dot{M}$ much less than the Eddington rate $\dot{M}_{\rm Edd}$ (*25*), corresponding to low-luminosity XRBs and AGN, observations primarily favor thick disks (*24*). The most compelling example is the supermassive black hole in our own galactic center, which is surprisingly dim ($L \approx 10^{36}$ erg s^{-1}, only ~100 times that of our Sun) and is a very inefficient accretor [$\varepsilon \approx 10^{-6}$ as estimated from the mass supply rate at large distances (*26*), although only a small fraction of this gas may actually reach the hole, as discussed below]. In systems such as this, the accretion is thought to take place via a hot two-temperature collisionless plasma in which the protons are much hotter than the electrons. The physics of such plasmas is not fully understood (*27*).

In the opposite limit of very high accretion rates, $\dot{M} \gg \dot{M}_{\rm Edd}$, the inflowing gas again does not radiate efficiently because there is so much accreting material that the radiation produced by the inflowing gas is trapped in the accretion flow and cannot escape before falling into the black hole (*28*). This is true until $\dot{M} \approx 10^{15}\ \dot{M}_{\rm Edd}$ (~1 $M_\odot$ s^{-1} for a solar-mass black hole), at which point the inflowing gas radiates away its gravitational binding energy via neutrinos rather than electromagnetic waves (*29*, *30*). At still higher accretion rates, even neutrinos are trapped in the inflowing gas and we again have thick disks. Such high accretion rates are never reached for black holes in XRBs or AGN, where characteristic rates are below the Eddington rate. They can, however, be achieved in the process of forming neutron stars and solar-mass black holes in the core collapse of massive stars (i.e., supernovae). Recent observations show that some gamma-ray bursts—explosions that release ~10^{51} ergs as gamma-rays in a few seconds—are associated with supernovae (*31*, *32*) and may be the signatures of newly formed black holes accreting the debris of their parent stars at ~1 $M_\odot$ s^{-1} and ejecting ultrarelativistic jets (*9*).

Finally, Fig. 2 illustrates a Goldilocks-like regime of accretion rates, $\dot{M} \approx \dot{M}_{\rm Edd}$, in which accretion proceeds via a thin disk that efficiently radiates away its gravitational potential energy. The precariousness of this regime has been highlighted by recent work showing that the fate of gas in an accretion flow depends critically on whether the disk is thin or thick. In thin disks, most of the mass supplied at large radii reaches the central black hole. By contrast, in thick disks, very little of the supplied mass ends up accreting into the hole. Instead, most of the mass circulates in convective motions (*33*, *34*) or is driven away in an unbound outflow (*35*, *36*). This in turn causes the amount of radiation from the accretion flow to decrease drastically. Because the fate of supplied matter depends so strongly on the mode of accretion (thin versus thick), it is likely that bright accreting black holes occupy a somewhat narrow range of accretion rates (Fig. 2).

In or Out?

The outflows from thick disks may be related to another outstanding observational and theoretical puzzle. Observations reveal that many accreting systems have powerful outflows of mass and energy. These are not unique to accreting black holes—young accreting stars show powerful outflows, as do some accreting neutron stars—but the outflows from black holes, particularly the collimated jets from massive black holes (*37*, *38*), are probably the most remarkable (Fig. 3).

Recent work—including direct numerical simulations of magnetized accreting plasma in General Relativity—has clarified the possible origin of these outflows. Simulations (Fig. 1) reveal a mildly relativistic outflow that emerges self-consistently from the accretion flow (*39*, *40*). The outflow is stronger and energetically more important for rapidly rotating black holes. Some of the

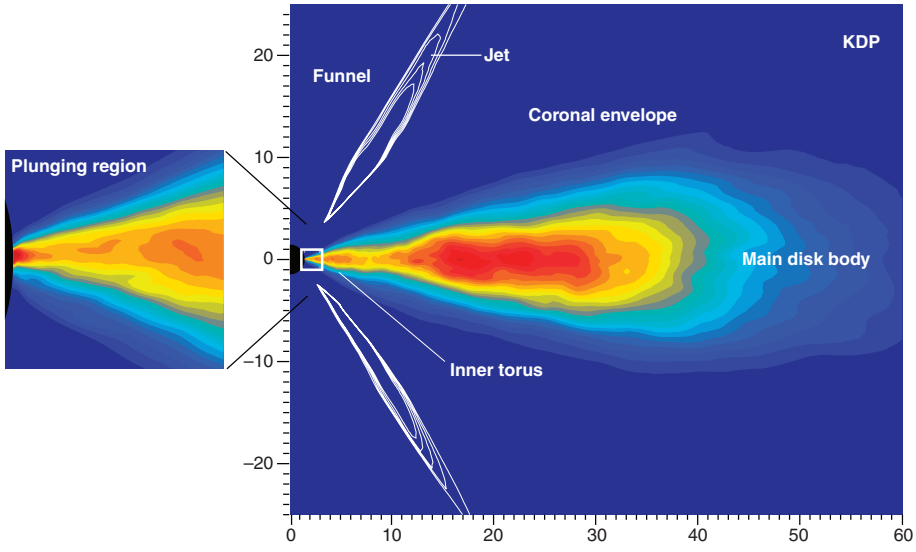


Fig. 1. The inner regions of an accretion disk around a black hole, as calculated in a General Relativistic magnetohydrodynamic numerical simulation (*39*). The black hole is at coordinates (0,0) with an event horizon of radius unity. The accretion disk rotates around the vertical direction (the axis of the nearly empty funnel region). Its density distribution is shown in cross section, with red representing the highest density and dark blue the lowest. Above the disk is a tenuous hot magnetized corona, and between the corona and the funnel is a region where there is ejection of mildly relativistic plasma, which may be related to the formation of jets. The example shown here corresponds to a radiatively inefficient thick disk. In the opposite limit of a thin disk, the vertical thickness of the disk would be much smaller, sometimes as small as 0.1% of the radius.

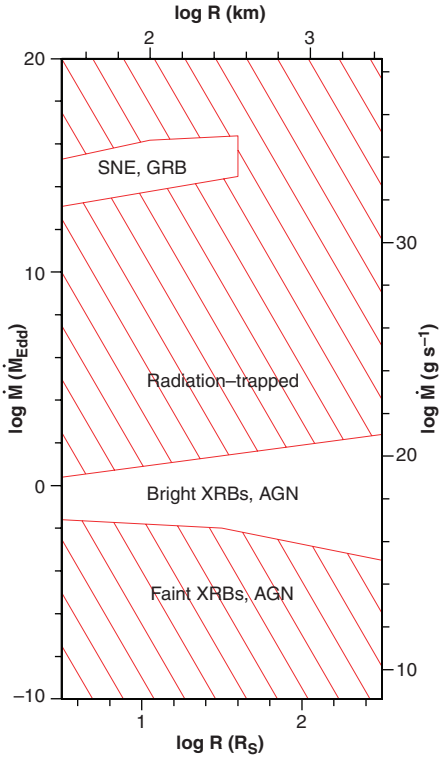


Fig. 2. Thin- and thick-disk regimes in the parameter space of accretion rate and radius. The figure covers 30 orders of magnitude in accretion rate to illustrate the broad applicability of this theoretical classification. The bottom horizontal axis is the radius of the accreting gas in units of the event horizon; the left vertical axis is the accretion rate in Eddington units (*25*). The top horizontal axis and right vertical axis give physical units for a 3 $M_\odot$ black hole. Shaded regions denote radiatively inefficient thick disks; nonshaded regions denote radiatively efficient thin disks. SNE, supernova explosions; GRB, gamma-ray bursts.

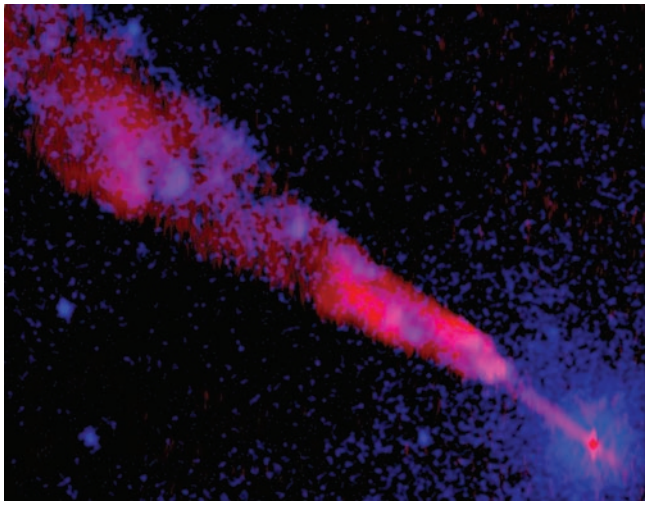


Fig. 3. Composite image of the inner 4000 light years of a jet in the galaxy Centaurus A (*66*). Blue shows x-ray emission as seen by the Chandra X-ray Observatory, and red shows radio emission as measured with the Very Large Array. The jet originates from the vicinity of a supermassive black hole at the center of the galaxy in the lower right of the image. The jet is straight and well-collimated, and has a total length much greater than that seen in the inner region highlighted here.

energy for this outflow comes directly from the rotational energy of the hole rather than from the accreting matter, a possibility that has been discussed for more than 30 years (*41*, *42*).

One physical process by which a jet can be produced by a spinning hole has been cleanly demonstrated in a numerical experiment (Fig. 4) (*43*–*45*). Just outside the horizon of a rotating black hole is a region of spacetime called the ergosphere where matter is not allowed to be at rest, but must rotate along with the hole. Because of this effect of frame dragging, magnetized plasma inside the ergosphere is twisted into corotation with the hole. The twist propagates out along the magnetic field lines as a torsional Alfvén wave, extracting the rotational energy of the black hole and driving an outflow. This is analogous to the mechanism by which outflows are driven from rotating magnetized stars (e.g., pulsars) or accretion disks (*46*, *47*).

Observationally, it is not clear whether the hole or the disk is the source of the outflows seen in accreting black holes (*48*, *49*). X-ray binaries show a strong correlation between the presence of a thick disk and the presence of radio emission from jets (*50*, *51*). There is related evidence in AGN as well, particularly for low-luminosity systems (*52*, *53*). This suggests that the dynamics of the underlying accretion disk is central to the production of jets, either because the outflow originates from the thick disk itself or because the magnetic field structure associated with the thick disk is more conducive to the extraction of energy from the central black hole.

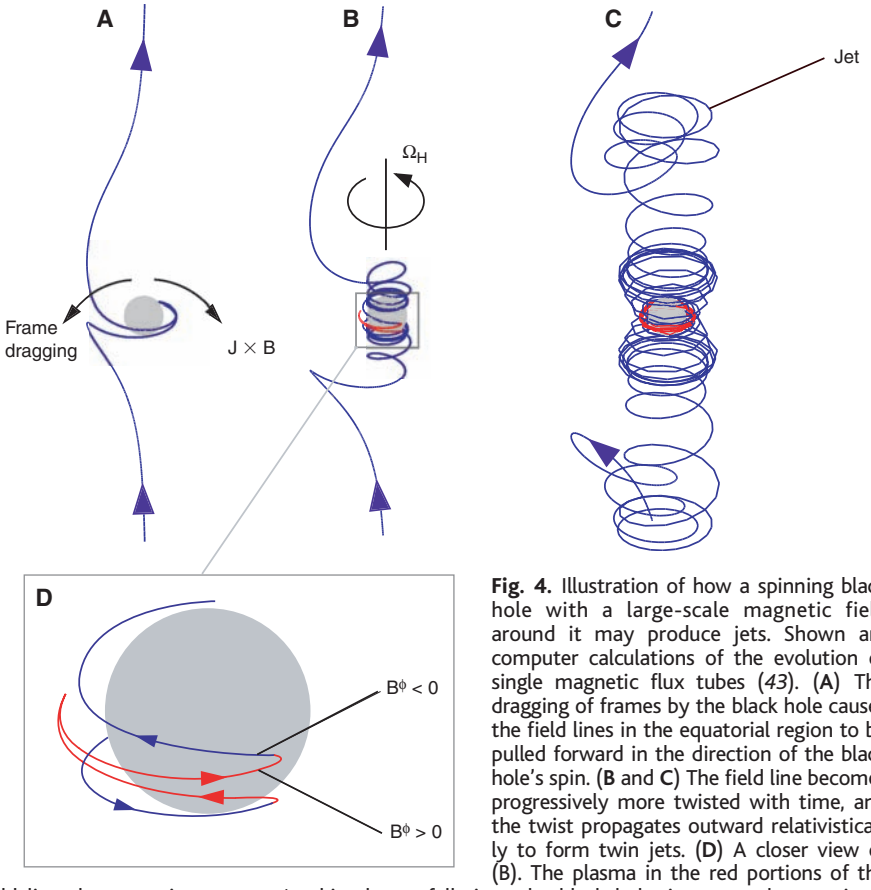


Fig. 4. Illustration of how a spinning black hole with a large-scale magnetic field around it may produce jets. Shown are computer calculations of the evolution of single magnetic flux tubes (*43*). (**A**) The dragging of frames by the black hole causes the field lines in the equatorial region to be pulled forward in the direction of the black hole's spin. (**B** and **C**) The field line becomes progressively more twisted with time, and the twist propagates outward relativistically to form twin jets. (**D**) A closer view of (**B**). The plasma in the red portions of the field lines has negative energy. As this plasma falls into the black hole, it causes the rotational energy of the hole to decrease. The extra energy comes out in the jet, in a process analogous to the original suggestion of Penrose (*41*) for extracting energy from a spinning black hole. **J**, electric current vector; **B**, magnetic field vector; B^ϕ, azimuthal component of **B**; Ω_H, spin vector of the black hole.

Spinning Holes

Numerical simulations of jets from black holes provide some support for the long-hypothesized connection between relativistic outflows and black hole spin. With the recent progress in measuring the masses of astrophysical black holes, spin is now the sole remaining property required for a full characterization of these objects. There is tantalizing evidence that most massive black holes in galactic nuclei rotate rapidly. The total mass in these black holes in the local universe can be compared with the total radiation produced by accretion over the history of the universe to infer the average efficiency ε of accreting black holes (*54*). This yields efficiencies $\varepsilon > 0.1$ (*55*, *56*), providing suggestive statistical evidence that most holes rotate appreciably, with the spin parameter $a*$ (*6*) being closer to unity than to zero.

Direct evidence for black hole spin in individual systems is, in principle, possible because spin imparts unique signatures on the structure of spacetime around black holes. Accretion disks around rotating holes extend inside the ergosphere, and the properties of radiation from gas in this inner region should reflect the spin of the hole. Some XRBs and AGN show lines from ionized iron in their x-ray emission. In several cases, the lines are remarkably broad, with a width comparable to the rest wavelength of the line (*57*, *58*). Although the interpretation and analysis are subtle, the line-broadening has been interpreted as being due to the extreme gravitational redshifts and Doppler shifts present in accretion disks around rotating black holes with $a* > 0.9$ (*59*).

Certain classes of XRBs—both those with black holes and those with neutron stars—show nearly periodic oscillations in their x-ray emission (*60*). These quasi-periodic oscillations (QPOs) have periods similar to the ro-

tation period of matter near the event horizon of the central black hole, which suggests that they arise as matter is just about to plunge through the horizon. Just as earthquakes can be used to study Earth's interior structure, QPOs may provide a means of mapping out the structure of spacetime around black holes, thereby constraining the spin of the hole (*61*).

These methods for estimating the spin are made possible because the gas in an accretion disk probes the relativistic region of spacetime near the black hole. An even more remarkable property of a black hole than the twisting of spacetime by its spin is the fact that it possesses an event horizon. Because the gas in an accretion disk flows into the hole through the horizon, it should, in principle, be possible to test whether black holes do have horizons. Several intriguing differences have been identified between the luminosities and spectra of XRBs with black holes and those with neutron stars. When combined with models of accretion, these differences are found to be consistent with black holes having event horizons and neutron stars having surfaces (*62*, *63*), thus confirming one of the most spectacular predictions of General Relativity.

Ideally, for all these studies, one would like to directly observe gas as it is about to plunge through the horizon of a black hole. In the near future, the most promising technique for doing so is very long baseline interferometry (VLBI), which uses radio telescopes on opposite sides of Earth to create the most detailed images of astronomical objects. At a wavelength of $\sim$0.3 mm, this technique can reach a resolution of $\sim 5 \times 10^{-11}$ rad $\approx 10^{-5}$ arc sec. This would be sufficient to see a coin the size of a U.S. quarter on the Moon. More relevant to observing gas near the horizon of a black hole is that the size of the horizon of the black hole in the galactic center is $\sim 10^{-5}$ arc sec on the sky. VLBI experiments are being planned to try to image the emission from gas close to this black hole (*64*), and there are exciting prospects for seeing General Relativistic effects in such observations, including signatures of black hole spin (*65*). This would represent a remarkable achievement for black hole astrophysics: from theorists' speculations to direct images in $\sim$100 years.

References and Notes

1. K. Freeman, J. Bland-Hawthorn, *Annu. Rev. Astron. Astrophys.* **40**, 487 (2002).
2. F. H. Shu, F. C. Adams, S. Lizano, *Annu. Rev. Astron. Astrophys.* **25**, 23 (1987).
3. P. Goldreich, Y. Lithwick, R. Sari, *Annu. Rev. Astron. Astrophys.* **42**, 549 (2004).
4. C. W. Misner, K. S. Thorne, J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
5. J. B. Hartle, *Gravity: An Introduction to Einstein's General Relativity* (Addison-Wesley, San Francisco, 2003).
6. The spin of a black hole is described by a dimensionless parameter a_*, defined such that the angular momentum of the hole is equal to $a_* GM^2/c$, where G is the gravitational constant, M is the mass of the hole, and c is the speed of light; a_* is restricted to the range 0 to 1. In addition to the two parameters, M and a_*, a black hole can in principle have a third parameter, electric charge, but astrophysical black holes are likely to be neutral.
7. C. M. Miller, E. J. M. Colbert, *Int. J. Mod. Phys. D* **13**, 1 (2004).
8. A. Heger, C. L. Fryer, S. E. Woosley, N. Langer, D. H. Hartmann, *Astrophys. J.* **591**, 288 (2003).
9. P. Mészáros, *Annu. Rev. Astron. Astrophys.* **40**, 137 (2002).
10. The bulge of a galaxy is a roughly spherical collection of old stars. In a spiral galaxy (like the Milky Way), the bulge dominates the inner region, whereas the more gas-rich disk and its younger stars extend farther out. In the case of an elliptical galaxy, the bulge constitutes almost all the stars in the galaxy.
11. J. Magorrian *et al.*, *Astron. J.* **115**, 2285 (1998).
12. L. Ferrarese, D. Merritt, *Astrophys. J.* **539**, L9 (2000).
13. K. Gebhardt *et al.*, *Astrophys. J.* **539**, L13 (2000).
14. R. Schödel *et al.*, *Nature* **419**, 694 (2002).
15. A. M. Ghez *et al.*, *Astrophys. J.*, in press (http://arxiv.org/abs/astro-ph/0306130).
16. J. Silk, M. J. Rees, *Astron. Astrophys.* **331**, L1 (1998).
17. A. King, *Astrophys. J.* **596**, L27 (2003).
18. N. Murray, E. Quataert, T. A. Thompson, *Astrophys. J.*, in press (http://arxiv.org/abs/astro-ph/0406070).
19. C. R. Kaiser, J. Binney, *Mon. Not. R. Astron. Soc.* **338**, 837 (2003).
20. A. C. Fabian *et al.*, *Mon. Not. R. Astron. Soc.* **344**, L43 (2003).
21. S. A. Balbus, J. F. Hawley, *Rev. Mod. Phys.* **70**, 1 (1998).
22. The physics of accretion in neutral disks is much less well understood.
23. J. Frank, A. King, D. J. Raine, *Accretion Power in Astrophysics* (Cambridge Univ. Press, Cambridge, 2000).
24. R. Narayan, R. Mahadevan, E. Quataert, in *Theory of Black Hole Accretion Disks*, M. A. Abramowicz, G. Bjornsson, J. E. Pringle, Eds. (Cambridge Univ. Press, Cambridge, 1998), pp. 148–182.
25. The Eddington luminosity is the luminosity at which the outward force due to radiation pressure becomes comparable to the inward force due to gravity. The mass accretion rate at which gas flowing into a black hole produces this luminosity is called the Eddington accretion rate, $\dot{M}_{\rm Edd}$. The numerical value of this quantity depends on the assumed radiative efficiency ε. For the traditional choice $\varepsilon = 0.1$, $\dot{M}_{\rm Edd} \approx 10^{18}(M_{\rm BH}/M_\odot)$ g s^{-1}.
26. F. Melia, H. Falcke, *Annu. Rev. Astron. Astrophys.* **39**, 309 (2001).
27. E. Quataert, W. Dorland, G. Hammett, *Astrophys. J.* **577**, 524 (2002).
28. M. A. Abramowicz, B. Czerny, J. P. Lasota, E. Szuszkiewicz, *Astrophys. J.* **332**, 646 (1988).
29. R. Popham, S. E. Woosley, C. Fryer, *Astrophys. J.* **518**, 356 (1999).
30. R. Narayan, T. Piran, P. Kumar, *Astrophys. J.* **557**, 949 (2001).
31. K. Z. Stanek *et al.*, *Astrophys. J.* **591**, L17 (2003).
32. J. Hjorth *et al.*, *Nature* **423**, 847 (2003).
33. R. Narayan, I. V. Igumenshchev, M. A. Abramowicz, *Astrophys. J.* **539**, 798 (2000).
34. E. Quataert, A. Gruzinov, *Astrophys. J.* **539**, 809 (2000).
35. R. D. Blandford, M. C. Begelman, *Mon. Not. R. Astron. Soc.* **303**, L1 (1999).
36. J. F. Hawley, S. A. Balbus, *Astrophys. J.* **573**, 738 (2002).
37. J. A. Zensus, *Annu. Rev. Astron. Astrophys.* **35**, 607 (1997).
38. I. F. Mirabel, L. F. Rodríguez, *Annu. Rev. Astron. Astrophys.* **37**, 409 (1999).
39. J.-P. de Villiers, J. F. Hawley, J. H. Krolik, *Astrophys. J.* **599**, 1238 (2003).
40. J. C. McKinney, C. F. Gammie, *Astrophys. J.* **611**, 977 (2004).
41. R. Penrose, *Riv. Nuovo Cim.* **1**, 252 (1969).
42. R. D. Blandford, R. L. Znajek, *Mon. Not. R. Astron. Soc.* **179**, 433 (1977).
43. V. Semenov, S. Dyadechkin, B. Punsly, *Science* **305**, 978 (2004).
44. S. Koide, *Astrophys. J.* **606**, L45 (2004).
45. S. S. Komissarov, *Mon. Not. R. Astron. Soc.* **350**, 1431 (2004).
46. P. Goldreich, W. H. Julian, *Astrophys. J.* **157**, 869 (1969).
47. R. D. Blandford, D. G. Payne, *Mon. Not. R. Astron. Soc.* **199**, 883 (1982).
48. P. Ghosh, M. A. Abramowicz, *Mon. Not. R. Astron. Soc.* **292**, 887 (1997).
49. M. Livio, G. I. Ogilvie, J. E. Pringle, *Astrophys. J.* **512**, 100 (1999).
50. E. Gallo, R. P. Fender, G. G. Pooley, *Mon. Not. R. Astron. Soc.* **344**, 60 (2003).
51. A. Merloni, S. Heinz, T. di Matteo, *Mon. Not. R. Astron. Soc.* **345**, 1057 (2003).
52. L. Maraschi, F. Tavecchio, *Astrophys. J.* **593**, 667 (2003).
53. N. M. Nagar, H. Falcke, A. S. Wilson, J. S. Ulvestad, *Astron. Astrophys.* **392**, 53 (2002).
54. A. Soltan, *Mon. Not. R. Astron. Soc.* **200**, 115 (1982).
55. M. Elvis, G. Risaliti, G. Zamorani, *Astrophys. J.* **565**, L75 (2002).
56. Q. Yu, S. Tremaine, *Mon. Not. R. Astron. Soc.* **335**, 965 (2002).
57. The rest wavelength of a line is the wavelength that would be measured in a frame of reference not moving or accelerating with respect to the source. Laboratory experiments on Earth measure this wavelength because the radiation source and the measuring device are fixed with respect to each other.
58. C. S. Reynolds, M. A. Nowak, *Phys. Rep.* **377**, 389 (2003).
59. C. S. Reynolds, L. W. Brenneman, D. Garofalo, in *From X-ray Binaries to Quasars: Black Hole Accretion on All Mass Scales*, T. Maccarone, R. Fender, L. Ho, Eds. (Kluwer, Amsterdam, in press) (http://arxiv.org/abs/astro-ph/0410116).
60. J. E. McClintock, R. A. Remillard, in *Compact Stellar X-ray Sources*, W. H. G. Lewin, M. van der Klis, Eds. (Cambridge Univ. Press, Cambridge, in press) (http://arxiv.org/abs/astro-ph/0306213).
61. S. Kato, *Publ. Astron. Soc. Jpn.* **53**, 1 (2001).
62. M. R. Garcia *et al.*, *Astrophys. J.* **553**, L47 (2001).
63. R. Narayan, *Astron. Geophys.* **44**, 6.22 (2003).
64. S. Doeleman, G. Bower, *Galactic Center Newsletter 18* (2004) (www.aoc.nrao.edu/~gcnews/home/gcnews.shtml).
65. H. Falcke, F. Melia, E. Agol, *Astrophys. J.* **528**, L13 (2000).
66. M. J. Hardcastle *et al.*, *Astrophys. J.* **593**, 169 (2003).

Expanded Repeat in Canine Epilepsy

Hannes Lohi,[1]* Edwin J. Young,[1]* Susan N. Fitzmaurice,[2] Clare Rusbridge,[3] Elayne M. Chan,[1] Mike Vervoort,[1] Julie Turnbull,[1] Xiao-Chu Zhao,[1] Leonarda Ianzano,[1] Andrew D. Paterson,[1] Nathan B. Sutter,[4] Elaine A. Ostrander,[4] Catherine André,[5] G. Diane Shelton,[6] Cameron A. Ackerley,[1] Stephen W. Scherer,[1] Berge A. Minassian[1]†

Epilepsy afflicts 1% of humans and 5% of dogs. More than 5% of purebred miniature wirehaired dachshunds (MWHDs) in the United Kingdom suffer an autosomal recessive progressive myoclonic epilepsy (PME), which we show to be Lafora disease (EPM2) (*1*), the severest form of teenage-onset epilepsy in humans (*2*). EPM2 is caused by mutations in the chromosome 6q24 *EPM2A* or 6p22 *EPM2B* genes (*3*). Using homozygosity and linkage analysis, we mapped the MWHD disease locus to canine chromosome 35 (*1*), which is syntenic in its entirety to human 6p21-25. We then cloned canine *Epm2b* (*1*). Polymerase chain reaction (PCR) failed in affected dogs across the 5′ half of the gene's single exon. The normal sequence in this region contains two consecutive identical dodecamers (D) and a third copy differing by a single nucleotide (T) (Fig. 1A). The corresponding region in other species is not repetitive and is shorter by 12 nucleotides, the length of one D repeat (Fig. 1A). Concerted modifications (*1*) of PCR conditions ultimately enabled sequencing across the repeat region in affected dogs and revealed bi-allelic expansion of the dodecamer repeat with 19 to 26 copies of the D sequence (Fig. 1B) (*1*). Comparing the amount of *Epm2b* mRNA in skeletal muscle from three affected dogs and two controls with quantitative reverse transcription–PCR (*1*) showed that affected mRNA levels were more than 900 times reduced (Fig. 1C).

To determine whether the extra D sequence is specific to MWHDs, we sequenced *Epm2b* from two normal unrelated dogs from each of 128 breeds. Sixty percent of their chromosomes had three repeats (2 D's and 1 T) and 40%, two repeats (1 D and 1 T). Almost all breeds had examples of both variants, in homozygous or heterozygous state. We tested the next non-MWHD PME case to present to the clinic, a basset hound, and found a homozygous 14-copy expansion of the repeat (Fig. 1B).

In the presence of the normal allele, PCR of the expanded allele was impossible. Deaminating carrier DNA cytosines (*1*, *4*) before PCR allowed amplification of the mutant allele and reliable carrier detection (Fig. 1B). In affected dog DNA, every C nucleotide in the expansion deaminated, ruling out expansion methylation as a mechanism for the absent mRNA. The actual mechanism may be adoption of RNA polymerase-obstructive secondary hairpin structures by the large GC-pure DNA expansion (*5*).

A single recurring cystatin-B promoter expansion causes most human PME (EPM1), rather than the more than 60 mutations in EPM2 genes. To our knowledge, EPM1 has never been described in dogs, likely because they do not have the cystatin-B dodecamer repeat. EPM2, on the other hand, is regularly reported (basset hounds, miniature and standard poodles, pointers, corgis, beagles, dachshunds, etc.) (*6*), likely because of recurrent *Epm2b* expansion events plus inbreeding.

The presence of the dodecamer repeat across canine breed barriers suggests that its origin predates dogs, and it might therefore be present in related species. Sequencing the *Epm2b* gene in 32 different carnivores (*1*) showed that the repeat is present and polymorphic across *Canidae* (wolves, dogs, foxes, coyotes, jackals, etc.) but not their closest relatives, the *Arctoidae* (bears, raccoons, otters, skunks, etc.). Whether the variable four–amino acid lengthening of the middle portion of malin, the *Epm2b* gene product (Fig. 1A), confers a property to this ubiquitin E3 ligase advantageous to canids remains to be seen. *Arctoidae* have a single copy of the D sequence (Fig. 1A) (*1*), indicating that its duplication occurred sometime between the canid-arctoid split 50 million years ago (Ma) and the appearance of extant canids 10 Ma. The D sequence is missing in *Felidae* (Fig. 1A) (*1*), suggesting it first appeared after the earlier feline divergence 60 Ma (*7*).

We have described a canine epilepsy mutation that represents a tandem repeat expansion outside humans and devised a test to detect and counteract it through controlled breeding. Affected animals outnumber human EPM2 patients and afford us valuable experience in treating this disease. Additional searches for large coding repeats may reveal other expansions central to inherited diseases.

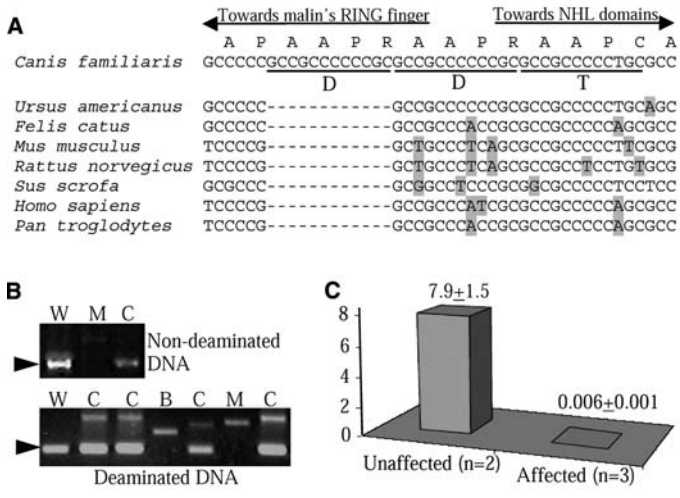


Fig. 1. (**A**) The canine *Epm2b* dodecamer repeat sequence and orthologs. (**B**) The expansion mutation. W, wild-type; M, affected MWHD; C, carrier; B, affected basset hound; arrowheads, normal alleles; other bands, mutant alleles. (**C**) Muscle *Epm2b* mRNA amounts normalized against *Gapdh* (*1*).

References and Notes

1. Materials and methods are available as supporting material on *Science* Online.
2. B. A. Minassian, *Adv. Neurol.* **89**, 199 (2002).
3. E. M. Chan *et al.*, *Neurology* **63**, 565 (2004).
4. A. Weinhaeusel *et al.*, *Hum. Mutat.* **22**, 404 (2003).
5. T. Saha, K. Usdin, *FEBS Lett.* **491**, 184 (2001).
6. T. Schoeman, J. Williams, E. van Wilpe, *J. Vet. Intern. Med.* **16**, 201 (2002).
7. C. M. Janis, K. M. Scott, L. L. Jacobs, *Evolution of Tertiary Mammals of North America* (Cambridge Univ. Press, Cambridge, 1998).

8. Supported by the Canadian Institutes of Health Research and Genome Canada.

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5706/81/DC1

Materials and Methods, SOM Text, Figs. S1 to S3, References and Notes, Movie S1

16 July 2004; accepted 18 October 2004
10.1126/science.1102832

[1]The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [2]Wey Referrals, Woking, Surrey GU21 5BP, UK. [3]Stone Lion Veterinary Centre, Wimbledon, London SW19 5AU, UK. [4]Fred Hutchinson Cancer Research Center, Seattle, WA 98109–1024, USA. [5]CNRS Génétique et Développement, 35043 Rennes, France. [6]University of California, San Diego, CA 92093, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: bminass@sickkids.ca

A Comprehensive Survey of the *Plasmodium* Life Cycle by Genomic, Transcriptomic, and Proteomic Analyses

Neil Hall,[1]*†‡ Marianna Karras,[2]† J. Dale Raine,[3]† Jane M. Carlton,[4,5,6] Taco W. A. Kooij,[2] Matthew Berriman,[1] Laurence Florens,[7]§ Christoph S. Janssen,[8] Arnab Pain,[1] Georges K. Christophides,[9] Keith James,[1] Kim Rutherford,[1] Barbara Harris,[1] David Harris,[1] Carol Churcher,[1] Michael A. Quail,[1] Doug Ormond,[1] Jon Doggett,[1] Holly E. Trueman,[3] Jacqui Mendoza,[3] Shelby L. Bidwell,[4] Marie-Adele Rajandream,[1] Daniel J. Carucci,[10] John R. Yates III,[7] Fotis C. Kafatos,[9] Chris J. Janse,[2] Bart Barrell,[1] C. Michael R. Turner,[8] Andrew P. Waters,[2]* Robert E. Sinden[3]*

Plasmodium berghei and *Plasmodium chabaudi* are widely used model malaria species. Comparison of their genomes, integrated with proteomic and microarray data, with the genomes of *Plasmodium falciparum* and *Plasmodium yoelii* revealed a conserved core of 4500 *Plasmodium* genes in the central regions of the 14 chromosomes and highlighted genes evolving rapidly because of stage-specific selective pressures. Four strategies for gene expression are apparent during the parasites' life cycle: (i) housekeeping; (ii) host-related; (iii) strategy-specific related to invasion, asexual replication, and sexual development; and (iv) stage-specific. We observed posttranscriptional gene silencing through translational repression of messenger RNA during sexual development, and a 47-base 3′ untranslated region motif is implicated in this process.

Rodent malaria parasite species provide model systems that allow issues to be addressed that are impossible with the human-infectious species *Plasmodium falciparum* and *P. vivax* (*1*). Three closely related species, *P. chabaudi*, *P. yoelii*, and *P. berghei*, are in common use in the laboratory. Comparative sequencing and analysis of the genomes of such model species, in addition to the complete genome sequence of *P. falciparum* (*2*), provide insights into the evolution of *Plasmodium* genes and gene families (*3*).

The malaria parasite differentiates into a series of morphologically distinct forms in the vertebrate and mosquito hosts. It alternates between morphologically related invasive stages (sporozoite, merozoite, and ookinete) and replicative stages (pre-erythrocytic, erythrocytic-schizont, and oocyst) interposed by a single phase of sexual development that mediates transmission from the human host to the anopheline vector (*1*). This report integrates genome sequence analyses of *P. berghei* and *P. chabaudi* with transcriptome and proteome data for *P. berghei*, allowing the categorization of protein expression, the analysis of regulation mechanisms for gene expression, and the identification of species-specific gene families and genes under selective pressure.

Genome sequencing and annotation. Partial shotgun sequencing (*4*) of the genomes of *P. c. chabaudi* (AS clone) and *P. berghei* (ANKA clone) generated assemblies of ~17 and ~18 Mb, respectively (Table 1 and table S1). Orthologous genes of these two genomes and of *P. y. yoelii* (*3*) and *P. falciparum* (*2*) were inferred through bidirectional BLAST searches (Table 2). Combining the gene predictions of the three rodent parasites revealed that 4391 genes had orthologs in *P. falciparum*. These orthologs represent a universal *Plasmodium* gene set (table S2), which was mainly distributed across the central "core" regions of the 14 *P. falciparum* chromosomes. For example, in the core region of *P. falciparum* chromosome 2, 144 of 158 genes had rodent parasite orthologs (Fig. 1), whereas in the subtelomeric regions, only 3 of 65 genes showed (low) homology to rodent parasite genes (figs. S1 to S14). In addition to BLAST analysis, we manually examined the orthology of gene models on the basis of the conservation of gene order between the rodent parasites and *P. falciparum*, resulting in the identification of an additional 109 orthologs (table S3). There were no orthologs in the rodent parasite genomes for 736 *P. falciparum* genes, and 161 of these were located in the core regions (table S3). The other 575 are located in the subtelomeric regions, and Markov clustering (*5*) of these *P. falciparum*–specific genes revealed that almost half could be assembled into 12 distinct gene families (Fig. 1 and table S4). Only five subtelomeric gene families are obviously shared between all the sequenced *Plasmodium* species (table S4) (*6*). Previous studies have shown that a subtelomeric gene family of *P. vivax*, the *P. vivax* interspersed repeats (*vir*) (*7*), has related gene families in *P. berghei* (*bir*), *P. chabaudi* (*cir*), and *P. yoelii* (*yir*) (*8*, *9*), and we suggest *pir* (*Plasmodium* interspersed repeats) to collectively describe the families. The *bir* and *cir* families code for highly variable proteins that share ~30% sequence identity at the amino acid level. The copy number appears to be much higher in *P. y. yoelii* (>800 copies) compared to *P. berghei* (180 copies) and *P. c. chabaudi* (138 copies).

Selective pressure. Comparison of orthologous genes of different species through

[1]Pathogen Sequencing Unit, The Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [2]Department of Parasitology, Malaria Group, Leiden University Medical Center, Netherlands. [3]Immunology and Infection Section, Department of Biological Sciences, Imperial College London, Sir Alexander Fleming Building, Imperial College Road, London SW7 2AZ, UK. [4]The Institute for Genomic Research (TIGR), 9712 Medical Center Drive, Rockville, MD 20850, USA. [5]Department of Pathobiology, College of Veterinary Medicine, University of Florida, Gainesville, FL 32608, USA. [6]Department of Molecular Microbiology and Immunology, Johns Hopkins University, Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205, USA. [7]Department of Cell Biology, The Scripps Research Institute, SR-11, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA. [8]Division of Infection and Immunity, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK. [9]European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 69117 Heidelberg, Germany. [10]Naval Medical Research Center, Malaria Program (IDD), 503 Robert Grant Avenue, Room 3A40, Silver Spring, MD 20910–7500, USA.

*To whom correspondence should be addressed. E-mail: nhall@tigr.org (N.H., genome), a.p.waters@lumc.nl (A.P.W., transcriptome), and r.sinden@imperial.ac.uk (R.E.S., proteome)

†These authors contributed equally to this work and are listed alphabetically. N.H. led the genome team; M.K., the transcriptome team; and J.D.R., the proteome team.

‡Present address: The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA.

§Present address: Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, MO 64110, USA.

models of nucleotide sequence evolution can be used to investigate variable (and positive or negative) selective pressures (*10*, *11*). We determined the relative number of synonymous (dS) versus nonsynonymous (dN) substitutions between orthologs of *P. berghei* and *P. chabaudi*. In general, we found that orthologous gene pairs are under purifying selection pressure (and have dN/dS < 1) and that the observed ratios of median values for genes of rodent parasites (Table 2) were similar to those reported for *Caenorhabditis elegans*/*C. briggsae* and mouse/human comparisons (*10*, *12*). This strong divergence from dN/dS = 1 suggests that most rodent malaria parasite gene models code for proteins and are not mispredictions or pseudogenes. The distribution of dN/dS ratios of genes containing transmembrane (TM) domains or signal peptides (SPs) (i.e., genes that may be extracellular) was greater than that of cytoplasmic proteins lacking these domains (Fig. 2A), indicating reduced purifying or increased diversifying pressure on SP/TM-containing proteins, possibly as a result of selective pressure from the host. When these data are correlated with expression data from the transcriptome and proteome analysis (table S5), we observe significant difference between the dN/dS values in SP/TM-containing and non–SP/TM-containing genes in blood-stage proteins but not vector-stage proteins (Fig. 2B). This indicates that diversifying selection might result from selective pressure from the host adaptive immune response, although some parasite proteins expressed in the vector are also clearly under diversifying selection. Annotated genes with the highest dN/dS values include many genes that one would expect to play a role in host-parasite interactions, such as reticulocyte binding protein (0.81), rhoptry-associated protein (0.94), and erythrocyte binding antigen (0.78). We have compared our data set with the data generated by a recent study of

Table 1. Genome summary statistics. A more detailed set of statistics is given in table S1.

Statistic	*P. berghei*	*P. c. chabaudi*	*P. y. yoelii*	*P. falciparum*
Size (bp)	17,996,878	16,866,661	23,125,449	22,853,764
No. contigs	7,497	10,679	5,687	93
Average contig size (bp)	2,400	1,580	4,066	213,586
Sequence coverage	4×	4×	5×	14.5×
No. protein coding genes	5,864*	5,698*	5,878	5,268

*An excessive number of gene models were predicted for *P. berghei* and *P. c. chabaudi* because of the fragmented nature of the genome sequence data for these species. Thus, the gene numbers indicated are for gene predictions where orthologs were identified in other *Plasmodium* species only.

Table 2. Genome comparisons between the four sequenced *Plasmodium* species. Av., average; *P.b.*, *P. berghei*; *P.c.*, *P. chabaudi*; *P.f.*, *P. falciparum; P.y., P. yoeli.*

Statistic	*P.b. vs. P.y.*	*P.b. vs. P.c.*	*P.c. vs. P.y.*	*P.y. vs. P.f.*	*P.b. vs. P.f.*	*P.c. vs. P.f.*
Av. protein identity (%)	88.2	83.2	84.6	61.2	62.9	61.9
Av. nucleotide identity (%)	91.3	87.1	88.1	69.6	70.3	70.1
Median dN	0.05	0.07	0.06	0.29	0.26	0.26
Median dS	0.026	0.49	0.53	49.4	26.1	26.5
Median dN/dS†	0.16	0.13	0.11	0.008	0.009	0.009
No. orthologous gene pairs	3153	4641*	3318	3375	3890	3842

*The high number of orthologs inferred between *P. c. chabaudi* and *P. berghei* compared to pairwise comparisons of the other species most likely reflects the method of automated annotation of both genomes, which used identical gene-finding algorithms (*4*). †Median dN/dS value represents the median value of dN/dS for every gene pair and is not calculated from the median dN and dS values for each comparison. The median dN/dS for comparisons with *P. falciparum* are low because of the saturation of synonymous changes in the alignments, resulting in high dS values.

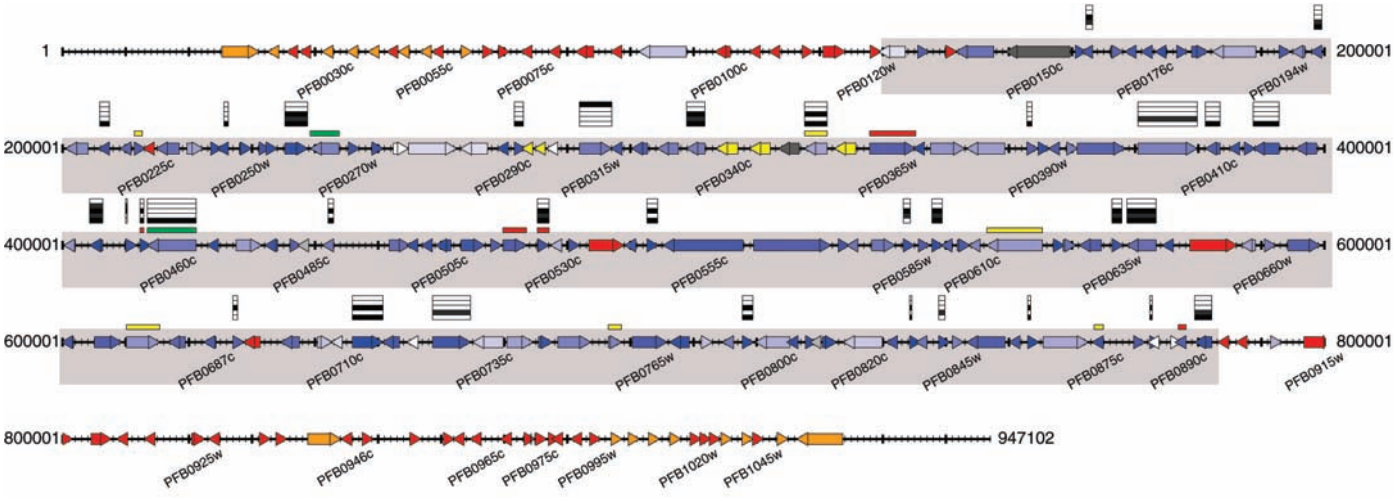
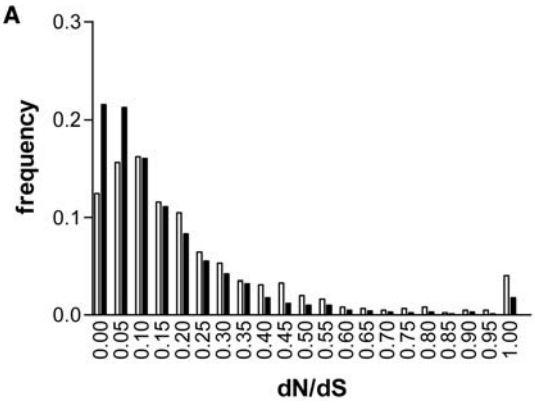


Fig. 1. Schematic map of *P. falciparum* chromosome 2. Arrow heads and boxes represent genes and their orientation on the DNA molecule. Thin and thick vertical lines represent 1-kb and 10-kb intervals, respectively. *P. falciparum* genes with orthologs in the rodent malaria parasite genomes are marked in shades of blue according to their degree of similarity, from light blue (indicating 40% identity) through dark blue (indicating 100% identity); white genes show <40% identity to their closest ortholog. Weak orthologs not detected by reciprocal BLAST analyses are indicated in dark gray and in light gray if the gene is absent in all rodent malaria parasite genomes. *P. falciparum* genes with no detectable ortholog are classified as follows: orange, *var*, *rif*, and *stevor* gene families; yellow, centrally located expanded gene families shared with the rodent malaria parasite; red, all other *P. falciparum* orphan genes. A full list of these genes and their classification can be found in table S3. Shaded areas of the map indicate the boundaries of the conserved chromosome core. Transcriptome and proteome data are marked above each gene where available. Transcripts that are up-regulated in asexual and gametocyte stages are shown as red or green horizontal lines, respectively; yellow lines denote genes that are up-regulated in both stages. Protein expression data are indicated by use of a bar code in which shading of each level indicates the following: top bar, sporozoite; second bar, oocyst; third bar, ookinete; fourth bar, gametocyte; and lowest bar, asexual stages. The identifier for every fifth gene (e.g., PFI0025c) is indicated. Schematic maps of all the *P. falciparum* chromosomes are shown in figs. S1 to S14.



B

Orthologs	SP or TM		Non-SP or TM		D	P
	n	median	n	median		
Genomics - all stages	726	0.174	3952	0.121	0.147	>0.001*
Proteome and transcriptome - all stages	147	0.157	501	0.100	0.217	>0.001*
- mammal expressed	106	0.159	311	0.098	0.237	>0.001*
- mosquito expressed	41	0.139	188	0.104	0.17	<0.05

Fig. 2. dN:dS ratios between pairs of orthologous genes in *P. berghei* and *P. c. chabaudi* and a comparison of genes containing SP or TM domains versus those lacking such domains. (**A**) Frequency distribution for all ortholog pairs. Open bars represent orthologs containing SP or TM domains; solid bars represent orthologs lacking such domains. (**B**) Analysis of distributions for all orthologs confirmed to be transcribed using transcriptome data or expressed using proteome data (table S2), partitioned according to their expression in mammalian or mosquito phases of the life cycle. The D variable represents the Kolmogorov-Smirnov test output statistic.

selection that measured codon volatility in *P. falciparum* (*13*). There are 15 *P. berghei* genes with a dN/dS ratio > 1 that have detectable orthologs in *P. falciparum*. Not all of these have scores indicating a high volatility, a result consistent with the facts that selection will be operating at different levels in different species and that volatility and dN/dS values measure selection over different time scales.

Gene expression. The asexual blood stage cycle of *P. berghei* takes 22 to 24 hours and gametocyte development 30 hours. Gametocytes are morphologically discernable from the asexual trophozoites only after 18 hours (fig. S18). Transcriptome data were obtained from three time points during the G1 phase (rings and young and mature trophozoites) and from two time points during the S/M phase (immature and mature schizonts), as well as from purified immature (24-hour) and mature (30-hour) gametocytes. The transcription profile of these stages was compared through a series of pairwise hybridizations to a *P. berghei* genome survey sequence (GSS) amplicon DNA microarray (*4*). Proteome data were collected from mixed asexual blood stages (containing both invasive and replicative stages), gametocytes during blood stage development, ookinetes, oocysts (days 9 to 12 postinfection), and salivary gland sporozoites and analyzed by multidimensional protein identification technology (*14*). The proteome analysis resulted in the identification of 1836 parasite proteins with high confidence (tables S6 to S8) and >5000 parasite proteins with relaxed filtering (*4*). By comparing expression data for the different life cycle stages, we could categorize proteins into the following four strategies of gene expression: (i) housekeeping, (ii) host-related expression, (iii) strategy-specific expression, and (iv) stage-specific expression.

Housekeeping. Of the 1836 proteins detected, 136 were expressed in at least four of the five stages analyzed (table S8). Given the lower number of proteins identified in the oocyst (277 proteins) and the sporozoite (134 proteins) compared to the other stages analyzed (733 to 1139 proteins), our analysis will have excluded some of the 301 proteins detected in asexual blood stages, gametocytes, and ookinetes (Fig. 3C). Recognizing that these 301 proteins were detected in both vertebrate and mosquito stages, we anticipate that some of these will also be expressed in oocysts and sporozoites.

Host-related expression. The proteome and transcriptome data sets revealed that enzymes of the tricarboxylic acid cycle, oxidative phosphorylation, and many other mitochondrial proteins were up-regulated in the gametocyte when compared to the asexual blood stages and were even more abundant in the ookinete (fig. S16 and table S8). These observations suggest that, as in trypanosomes (*15*), mitochondrial activity increases in the gametocyte as a preadaptation to life in the mosquito vector, and are consistent with the more complex organization of mitochondria in gametocytes (*1*, *16*). Mitochondrial activity apparently continues to increase in the ookinete.

Strategy-specific expression. Strategy-specific protein expression is related to invasion, asexual replication, or sexual development. We uniquely detected 966 proteins in invasive zoite (merozoite, ookinete, or sporozoite)–containing preparations, of which 234 were shared between at least two of the three invasive stages but not with the replicative or sexual stages (Fig. 3A). Gliding motility typifies the invasive stages of apicomplexans, and many proteins with a (putative) role in this process were detected. Micronemes and rhoptries are secretory organelles specific to the invasive stages. Although 10 known rhoptry proteins were detected in blood stages and sporozoites, these rhoptry proteins were absent from ookinetes. In contrast, most known micronemal protein families were detected in all zoites but with clear stage-specific expression of different family members. Perforin-like proteins, first described in the micronemes of *P. yoelii* sporozoites (*17*), contain a membrane attack complex/perforin (MACPF)–like domain and were found both in ookinetes and sporozoites but not in merozoites. We suggest a role for these molecules in parasite entry to and/or egress from target cells, given the role of MACPF-like domains in the formation of pores. Both the ookinete and sporozoite can traverse through several host cells (*18*, *19*), whereas a merozoite enters a target cell only once. Our data therefore support the concept that microneme proteins mediate motility and disruption of the host cell plasma membrane and that the rhoptry proteins are essential to genesis of the parasitophorous vacuole and host cell survival.

We uniquely detected 472 proteins in replicative stages, i.e., blood stages and oocysts (Fig. 3B). Not unexpectedly and as consistent with findings in *P. falciparum* (*20*–*22*), the majority of these genes encode proteins involved in cell growth or division, DNA replication, transcription, translation, and protein metabolism. The more detailed transcriptome analysis of blood-stage gene expression confirmed a cell cycle–related timing of transcription of these genes during the G1 and S/M phases (figs. S18 and S19) and revealed that 215 and 355 were upregulated in the G1 and the S/M phases, respectively.

During the first 18 hours of development, gametocytes and asexual trophozoites share the same features of the G1 phase of growth. Subsequently, the gametocytes differentiate into either males that prepare for DNA replication and mitosis or females that prepare for postzygotic growth. Transcriptome analysis demonstrated that 58% of the G1 proteins (125 genes) and 59.4% of the S/M proteins (199 genes) were also upregulated in gametocytes (fig. S19), and the proteome data also emphasized the similarity between protein expression in asexual blood

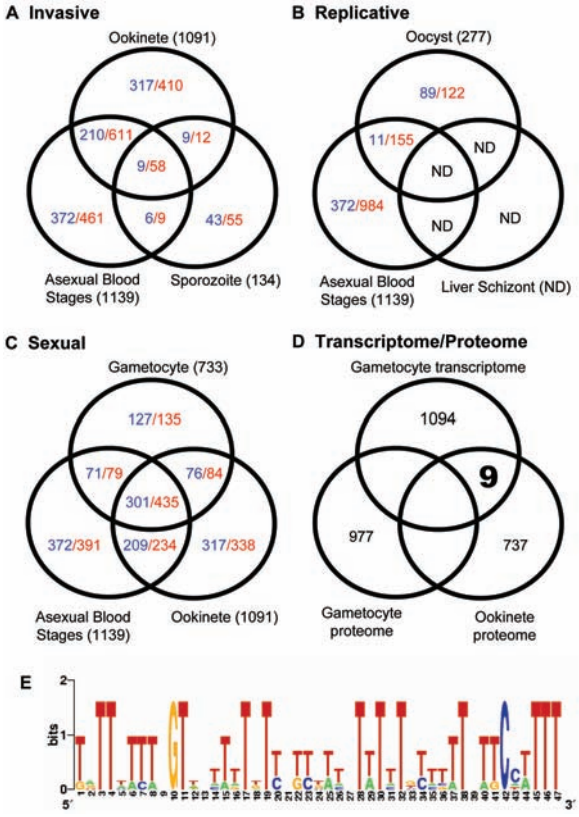


Fig. 3. Different strategies of protein and gene expression during the malaria life cycle. Venn diagrams illustrate the overlap in proteins detected in the life stages involved in (**A**) invasion, (**B**) replication, and (**C**) sexual development. The total number of proteins detected in each stage is shown in parentheses. Blue numbers represent proteins detected exclusively in the stages shown; red numbers represent proteins detected in the combination of stages shown out of the three stages included in each of the Venn diagrams (i.e., these proteins could also be shared with stages not shown in the figure). ND, not done. (**D**) A Venn diagram representing the comparison of the gametocyte transcriptome with the proteomes of the gametocyte and the ookinete. The numbers indicate the individual transcripts and proteins in each analysis. The number of gametocyte proteins includes proteins identified in *P. berghei* during this study and proteins identified from *P. falciparum* gametocytes (*21*). The bold number 9 in the intersect indicates the number of gametocyte transcripts found exclusively as ookinete proteins as a result of this study. (**E**) A WebLogo (*29*) representation of the 47-base motif found within 500 base pairs (bp) downstream of the open reading frames of six of the nine implicated translationally repressed transcripts for which 3′UTR sequence was available. The point size of the letter is proportional to the frequency of the appearance of each nucleotide at each position.

stages and gametocytes (514 proteins were shared between these stages) (table S8). Despite these similarities, the described unique morphologies indicate that sexual development is a fundamental developmental switch. This is shown by the specific up-regulation of transcription of 977 genes (*4*) (table S10 and fig. S19), including many of the known gametocyte-specific genes, and by the detection of 127 unique proteins in the proteome (Fig. 3C).

Stage-specific expression. Just over half (948) of the proteins detected in the proteome analysis were found in one stage only, suggesting that stage-specific specialization is substantial. However, many of these stage-specific proteins belong to protein families whose expression is strategy-specific, reflecting both conserved mechanisms of parasite development between different stages and subtle molecular adaptations dictated by specific parasite-host interactions. For example, gene families encoding proteins that contain MACPF-like or TSP/vWA domains are examples of strategy (invasion)–specific expression whose members are stage-specifically expressed. Unexpectedly, the PIR family belongs to this category: Members of the BIR protein family were detected in all stages, but 92% were exclusive to a single stage (fig. S15 and table S8). Peptides were found matching 34 of ~180 predicted *P. berghei* genes, and transcription of *bir* genes was detected in both the asexual blood stage and gametocytes (tables S9 and S10). Although *pir* are thought to play a role in immune evasion of the blood stages by antigenic variation (*7*), ~9% of the total BIR repertoire in our analysis was expressed only in the mosquito stages, suggesting that these proteins may have other key functions.

Posttranscriptional gene silencing (PTGS). It has been proposed that transcripts in *Plasmodium* are essentially produced when needed (*22*), the so-called ‘‘transcripts to go’’ model (*23*). However, it has been established that the abundant transcripts for *P28* in developing and mature female gametocytes are in a state of translational repression (TR) (*24*), one mechanism by which PTGS is exercised. In addition, RNA binding proteins (Puf proteins) (*25*) that play a role in TR are found in *Plasmodium* and are specifically up-regulated in gametocytes and sporozoites (*20*, *26*). Therefore, we compared the gametocyte transcriptome with the proteomes of both gametocytes and ookinetes to determine if additional gametocyte-specific transcripts might be subject to TR. Nine new genes were identified for which transcripts were detected in gametocytes but with protein products specific to the ookinete stage (Fig. 3D and table S11). The analysis of the 3′ untranslated regions (UTRs) of seven of these genes (for two genes, there was insufficient 3′UTR sequence for analysis) and the 3′UTRs of *Pbs28* and *Pbs25* by the motif identifier program MEME (*27*) revealed a 47-base motif found in six of the analyzed sequences within 1 kb of the 3′ end of the stop codon (Fig. 3E and fig. S17) (*E* value = $4.8e^{+002}$). Puf proteins bind to a UUGU motif in 3′UTR regions (*25*, *28*), and the 3′UTR regions of all seven candidates and *Pbs28* were enriched for this motif ($P \leq 0.001$), which was found as a submotif in the 47-base motif. The 47-base motif was used to search the entire *P. berghei* genome database with MAST (*27*), and 20 additional genes were identified that had the same motif within 1 kb of their 3′UTR ($E < e^{-05}$), giving a total of 29 TR candidates. Of these, 22 had orthologs in *P. falciparum*. Eighteen are up-regulated in gametocytes (16 genes) and/or sporozoites (5 genes), but only two were observed in gametocyte proteomes (table S11). Analysis of 1 kb downstream of the stop codon of 20 of these *P. falciparum* orthologs, including *pfs25* and *pfs28*, failed to identify a sequence analogous to the *P. berghei* motif. Nevertheless, visual inspection identified numerous UUGU motifs at analogous positions. This lack of sequence similarity of the predicted 3′UTR binding motif is consistent with the significant sequence diversity in the predicted gene models of the *Puf* orthologs of *P. falciparum* and *P. y. yoelii* (*25*). The paucity of annotated transcription factors (*2*, *28*) and the phased expression of blood-stage transcripts have led to the proposal that PTGS is a major mechanism of the regulation of gene expression in *Plasmodium* (*28*). Our data suggest that, at least in the gametocyte and possibly the sporozoite, TR may be an important component of these regulatory mechanisms.

The integration and initial analysis of the four data sets presented here has permitted insights concerning genome evolution, expression of multigene families, and mechanisms of posttranscriptional gene regulation in rodent malaria parasites. This initial overview will be developed further and, as demonstrated here, will continue to emphasize the value of model systems for the study of orthologous features of human malaria parasites.

References and Notes

1. R. E. Sinden, in *Rodent Malaria*, R. Killick-Kendrick, W. Peters, Eds. (Academic Press, London, 1978), pp. 85–168.
2. M. J. Gardner *et al.*, *Nature* **419**, 498 (2002).
3. J. M. Carlton *et al.*, *Nature* **419**, 512 (2002).

4. Materials and methods are available as supporting material on *Science* Online.
5. A. J. Enright, S. Van Dongen, C. A. Ouzounis, *Nucleic Acids Res.* **30**, 1575 (2002).
6. T. Y. Sam-Yellowe *et al.*, *Genome Res.* **14**, 1052 (2004).
7. H. A. del Portillo *et al.*, *Nature* **410**, 839 (2001).
8. K. Fischer *et al.*, *Mol. Microbiol.* **48**, 1209 (2003).
9. C. S. Janssen, M. P. Barrett, C. M. Turner, R. S. Phillips, *Proc. R. Soc. London Ser. B* **269**, 431 (2002).
10. L. D. Stein *et al.*, *PLoS Biol.* **1**, E45 (2003).
11. T. Endo, K. Ikeo, T. Gojobori, *Mol. Biol. Evol.* **13**, 685 (1996).
12. R. H. Waterston *et al.*, *Nature* **420**, 520 (2002).
13. Y. S. Han, J. Thompson, F. C. Kafatos, C. Barillas-Mury, *EMBO J.* **19**, 6030 (2000).
14. M. P. Washburn, D. Wolters, J. R. Yates III, *Nature Biotechnol.* **19**, 242 (2001).
15. K. Vickerman, *Nature* **208**, 762 (1965).
16. J. Krungkrai, P. Prapunwattana, S. R. Krungkrai, *Parasite* **7**, 19 (2000).
17. K. Kaiser *et al.*, *Mol. Biochem. Parasitol.* **133**, 15 (2004).
18. J. B. Plotkin, J. Dushoff, H. B. Fraser, *Nature* **428**, 942 (2004).
19. M. M. Mota, J. C. Hafalla, A. Rodriguez, *Nature Med.* **8**, 1318 (2002).
20. K. G. Le Roch *et al.*, *Science* **301**, 1503 (2003).
21. L. Florens *et al.*, *Nature* **419**, 520 (2002).
22. Z. Bozdech *et al.*, *PLoS Biol.* **1**, E5 (2003).
23. A. P. Waters, *Science* **301**, 1487 (2003).
24. M. G. Paton *et al.*, *Mol. Biochem. Parasitol.* **59**, 263 (1993).
25. L. Cui, Q. Fan, J. Li, *Nucleic Acids Res.* **30**, 4607 (2002).
26. R. E. Hayward *et al.*, *Mol. Microbiol.* **35**, 6 (2000).
27. T. L. Bailey, M. Gribskov, *J. Comput. Biol.* **5**, 211 (1998).
28. R. M. Coulson, N. Hall, C. A. Ouzounis, *Genome Res.* **14**, 1548 (2004).
29. G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, *Genome Res.* **14**, 1188 (2004).
30. We thank J. Dame (University of Florida) for the gift of the *P. berghei* GSS library, J. Langhorne (National Institute of Medical Research) for providing *P. c. chabaudi* DNA, R. G. Sadygov (The Scripps Research Institute) for expert computer programming, G. Butcher (Imperial College London) and M. Gardner (TIGR) for helpful advice with this manuscript, and M. Aslett and T. Kulikova for assisting with the database submission. Supported by the Wellcome Trust, the European Union, the Office of Naval Research, the U.S. Army Medical Research and Materiel Command, and the NIH. J.D.R. and J.M. are funded by the Wellcome Trust, M.K. by EU grant nos. RTN1-1999-00008 and QLK2-CT-1999-00753 and grant no. 050-10-053 from the Nederlands organisatie voor Wetenschappelijk Onderzoek, and H.E.T. by the EU MALTRANS consortium. The sequences have been submitted to EMBL under the accession prefixes CAAI for *P. berghei* and CAAJ for *P. chabaudi*. All data sets are available through the *Plasmodium* genome sequence data repository, PlasmoDB, at www.plasmodb.org and genome annotation at www.genedb.org.

Supporting Online Material
www.sciencemag.org/cgi/content/full/307/5706/82/DC1
Materials and Methods
Figs. S1 to S19
Tables S1 to S11
References and Notes

6 August 2004; accepted 2 November 2004
10.1126/science.1103717

REPORTS

Gigantic Photoresponse in ¼-Filled-Band Organic Salt $(EDO\text{-}TTF)_2PF_6$

Matthieu Chollet,[1] Laurent Guerin,[1,2] Naoki Uchida,[1] Souichi Fukaya,[1] Hiroaki Shimoda,[1] Tadahiko Ishikawa,[1] Kazunari Matsuda,[3] Takumi Hasegawa,[4] Akira Ota,[5] Hideki Yamochi,[6,8] Gunzi Saito,[5] Ryoko Tazaki,[7,8,9] Shin-ichi Adachi,[7,8] Shin-ya Koshihara[1,7,8]*

We report that the organic salt $(EDO\text{-}TTF)_2PF_6$ with ¾-filled-band (¼-filled in terms of holes), which forms an organic metal with strong electron and lattice correlation, shows a highly sensitive response to photoexcitation. An ultrafast, photoinduced phase transition from the insulator phase to the metal phase can be induced with very weak excitation intensity at near room temperature. This response makes the material attractive for applications in switching devices with room-temperature operation. The observed photoinduced spectroscopic change shows that this photoinduced phase transition process is caused by the cooperative melting of charge ordering assisted by coherent phonon generation.

Organic charge transfer (CT) A_2B salts composed of a cation (or anion) "A" and a counter ionic "B" in the ratio of 2:1 can display a variety of electronic and magnetic properties, such as superconductivity (*1*, *2*), metal-insulator (M-I) transition (*1*–*3*), magnetic frustration (*4*), ferroelectricity (*5*), and even magneto-dielectric coupled behaviors (*6*). The appearance of charge ordering (CO) or a Mott transition accompanied with dimeric distortion in a ¼-filled or a ¾-filled (¼-filled in terms of holes) network of A molecules with a one-dimensional (1D) or 2D structure plays a key role as the basis for these exotic natures (*1*, *2*). Recent theoretical studies have revealed that the critical balance among physical parameters, such as bandwidth (W), onsite Coulomb interaction (U), and nearest-neighbor Coulomb interaction (V), leads to the appearance of such novel ground states (*7*–*10*).

These previous studies and a report on the highly efficient photocarrier generation in one of the A_2B salts $(DCNQI)_2Cu$ (*11*) have stimulated research on the photoinduced phase transition (PIPT) in A_2B salts. In a crystal that shows multi-instability of its free energy due to critical balance among intrinsic cooperative interactions, the macroscopic phase transition accompanied with large changes in electronic, magnetic, and lattice structures may be triggered by weak photoexcitation (*12*, *13*), because even a low density of photoexcited species can affect and switch the cooperative interaction (*12*–*14*). From the viewpoint of the highly efficient and nonlinear amplification of the response to the weak photoexcitation via a cooperative channel in a condensed system, this exotic photo effect called PIPT is analogous to the "domino effect" on the molecular scale.

Here, we report that a quasi-1D, ¼-filled (in terms of holes) A_2B salt $(EDO\text{-}TTF)_2PF_6$ shows highly sensitive and ultrafast PIPT from an insulator (I) phase accompanied with a CO to a metal (M) phase up to about room temperature (~265 K). [EDO-TTF

[1]Department of Materials Science, Tokyo Institute of Technology, 2-12-1 Oh-okayama, Meguro-ku, Tokyo 152-8551, Japan. [2]Groupe Matière Condensée et Matériaux, Centre National de la Recherche Scientifique, Unite Mixte de Recherche-6626, Université de Rennes 1, Rennes, France. [3]Kanagawa Academy of Science and Technology and Japan Science and Technology Agency (JST), Precursory Research for Embryonic Science and Technology (PRESTO), 3-2-1, Sakado, Takatsu-ku, Kawasaki 213-0012, Japan. [4]Department of Physics, Tokyo Institute of Technology, Japan. [5]Division of Chemistry, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan. [6]Research Center for Low Temperature and Materials Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan. [7]High Energy Accelerator Research Organization (KEK) and [8]JST, Exploratory Research for Advanced Technology (ERATO), 1-1 O-ho, Tsukuba 305-0801, Japan. [9]Graduate School of Science and Technology, Chiba University 1-33, Yayoi-cho, Inage-ku, Chiba 263-8522, Japan.

*To whom correspondence should be addressed. E-mail: skoshi@cms.titech.ac.jp

(ethylenedioxytetrathiafulvalene) acts as an electron donor (D).] The origin of this exotic effect has been assigned to the photoinduced cooperative and nonequilibrium melting of a CO, based on time-resolved spectroscopic measurements. A dynamic study has revealed that the coherent phonon generation via electron-lattice (E-L) interaction seems to be strongly coupled with the observed highly efficient, so-called gigantic, photoinduced metallization process completed within 1.5 ps. In addition, the speed of recovery from the photoinduced M to the original I phase can be widely controlled by changing the excitation intensity (1 ps to 100 μs).

Recently, a quasi-1D, ¼-filled A_2B salt $(EDO\text{-}TTF)_2PF_6$ has been developed to realize an organic metal with strong E-L interaction that results from molecular deformations on a subnanometer scale (*15*). As expected, this crystal shows an exotic M-I phase transition that arises from CO, accompanied by a large structural change based on molecular deformation of EDO-TTF and a doubling of unit cells at room temperature (T_c = 278 K) (Fig. 1) (*15*–*18*). This M-I transition in $(EDO\text{-}TTF)_2PF_6$ can be sensitively probed spectroscopically, as plotted in Fig. 2A. The reflection band observed around 1.3 eV in the low-temperature I phase has been assigned to CT excitation among EDO-TTF (D) molecules accompanied by CO, i.e., excitation from D^+D^+ to $D^{2+}D^0$, directly reflecting the appearance of CO (*17*, *18*).

We prepared a crystal of $(EDO\text{-}TTF)_2PF_6$ with a surface of 0.1 mm by 0.2 mm using a previously reported method (*15*). We excited the crystal by laser light with a pulse width of 0.12 ps, and the excitation photon energy (1.55 eV) was nearly resonant to this CT band. The reflection spectra 10 ps before (Δt = –10 ps) and 3 ps after (Δt = 3 ps) the photoexcitation (Fig. 2B) were observed at 180 K and 265 K, with a resolution time of 0.25 ps. A similarly large and fast reflectivity change triggered by photoexcitation was confirmed in a wide temperature range below 265 K, and this change disappeared above T_c. The obtained spectral changes indicate that the intensity of the CT (D^+D^+ to $D^{2+}D^0$) band dramatically decreased and, in contrast, a rather weak and broad band, which seems to be characteristic of the M phase, appeared just after photoexcitation. Thus, CO melting, accompanied by I-M phase conversion, occurs within 3 ps after excitation by 0.12 ps light pulse. Movie S1 shows that the photoinduced modulation in reflectivity is large enough to be directly observed with an ordinary charge-coupled device camera after excitation with a white probe light 0.2 ps in width.

Compared with the reflectivity spectra of I and M phases under conditions of thermal equilibrium, the estimated photoinduced I-M conversion efficiency based on the reflectiv-

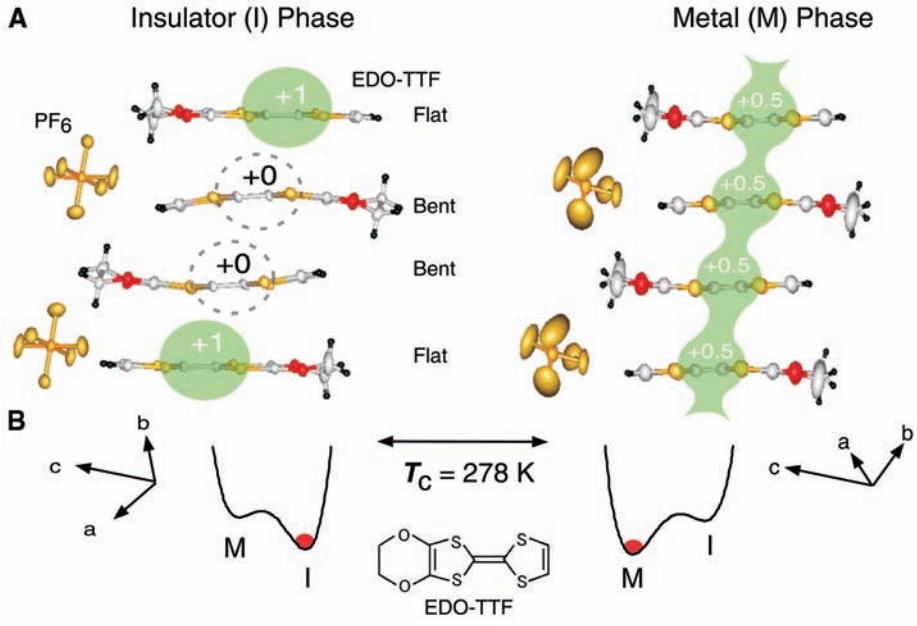


Fig. 1. (**A**) Schematic views of the lattice and electronic structural changes accompanying the M-I phase transition in $(EDO\text{-}TTF)_2PF_6$. A side view of an EDO-TTF molecule is shown. The unit cell includes two and four EDO-TTF molecules in M and I phases, respectively (*15*). In the I phase, holes are localized on EDO-TTF molecules with a flat structure due to CO, and quasi-neutral molecules show a bent structure. In the M phase, charges (holes) are delocalized and PF_6 (acceptor) molecules exhibit disorder (*15*–*18*). (**B**) Schematics for free-energy change accompanying M-I transition and the structure of the EDO-TTF molecule.

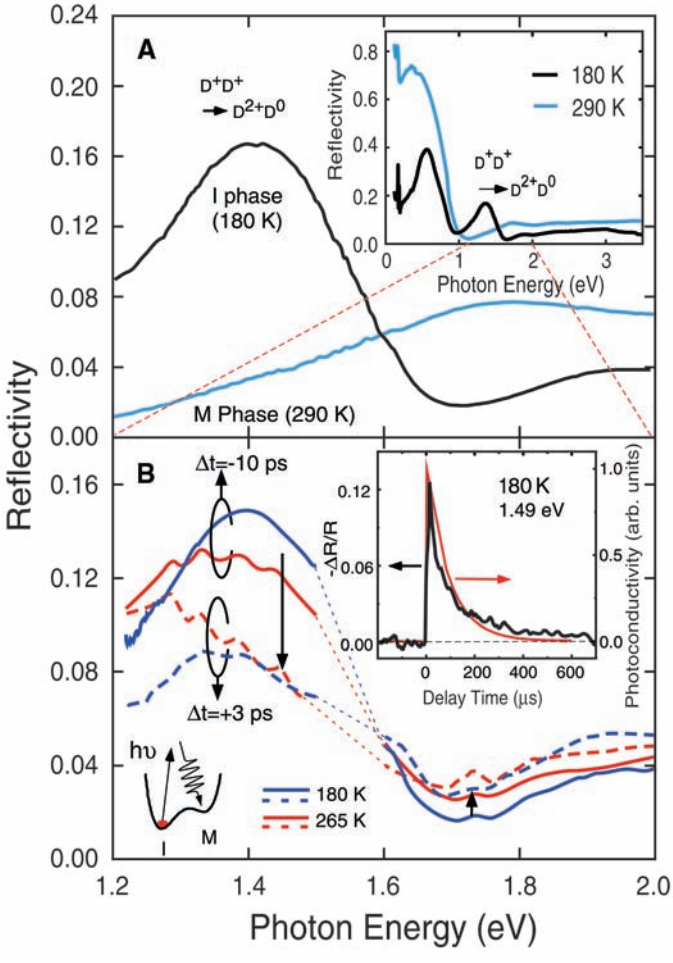


Fig. 2. (**A**) Reflectivity spectra for the M (blue line) and I (black line) phases observed at 290 and 180 K, respectively. (Inset) Reflectivity spectra in the wide photon energy region for the M (blue line) and I (black line) phases, respectively. (**B**) Reflectivity spectra observed at 10 ps before (Δt = –10 ps, solid lines) and 3 ps after (Δt = +3 ps, dashed lines) photoexcitation at the sample temperatures of 180 K (blue lines) and 265 K (red lines). The excitation density was 6×10^{18} cm^{-3} (*19*). The photoreflectivity spectra in the region between 1.5 and 1.6 eV could not be observed because of the strong scattering of the pump laser light. (Inset) Time profiles for $\Delta R/R$ at 1.49 eV (black line) and photoconductivity (red line) in the μs-ms region observed at 180 K.

ity change at Δt = 3 ps reached about 50% with an excitation intensity of 6 × 10^{18} photons cm^{-3}, corresponding to a single excitation photon for every 500 donor molecules (*19*). Such highly efficient conversion indicates a strong cooperative effect in this crystal, as expected from strong E-L coupling.

The inset in Fig. 2B shows the photoinduced reflectivity change ($\Delta R/R$) probed with 1.49 eV light and the photoconductivity at 180 K in the microsecond region with an excitation density of 6 × 10^{18} cm^{-3}. Both signals relaxed back to the original value within 300 to 400 μs. Thus, we could use light excitation with a 1 kHz repetition rate for a pump-probe experiment. The electric conductivity was largely enhanced (more than five orders of magnitude) just after photoexcitation within the resolution time (1 μs), which was consistent with the idea that the observed photoinduced effect arises from the I-M phase conversion.

A time profile of the spectral changes was observed at 1.38 and 1.72 eV with an excitation density of 6 × 10^{18} cm^{-3} (Fig. 3A). For Δt > 1.5 ps, the $\Delta R/R$ value remained constant, and the spectral shape was consistent with the I-M transition as a result of CO melting; thus, it can safely be concluded that the phase conversion process was completed within 1.5 ps for a wide temperature region. Such an ultrafast photoconversion process cannot be explained simply by the light-induced heating effect. In addition, as indicated by triangles in Fig. 3A, a vibrational structure with a large magnitude was observed. The period of this vibration, estimated from the interval between the first and second peaks [about 0.5 ps (70 cm^{-1}) at 180 K] has little dependence on the probe photon energy. Thus, this effect cannot be simply explained by strain-wave propagation induced by light excitation coupling with an acoustic mode as discussed previously (*20*–*22*).

As shown in the inset of Fig. 3A, several Raman modes in the low-energy region were observed for the I-phase crystal. The 84 cm^{-1} band observed at 180 K (red) softened as the sample temperature was increased (Fig. 3B, inset, red circles) (*23*). The same behavior has been observed in the temperature dependence of the vibration interval estimated from the time profiles of $\Delta R/R$ at various temperatures (Fig. 3B). The results are summarized and indicated in the inset of Fig. 3B by black squares. In addition, this vibrational mode was also Raman active in M phases (*23*). Based on these data, it is reasonable to attribute observed dynamic vibration in the time domain to the coherent phonon-like vibration that corresponds to this Raman-active optical mode, which exists even in the photoproduced M phase.

The present result indicates that the E-L interaction via the optical phonon mode plays important roles in driving the highly sensitive metallization process and stabilizing the photoinjected M domain. The vibrational mode calculation for a single EDO molecule in ionic and neutral states suggests the existence of several modes around a frequency region of 100 cm^{-1}, which can be attributed to a bending motion of the molecular plane (*23*). This result supports the idea that the observed coherent vibration in the $\Delta R/R$ time profile is due to strong coupling between CO and distortion of the molecular plane, which plays an essential role in thermally induced I-M transition in (EDO-

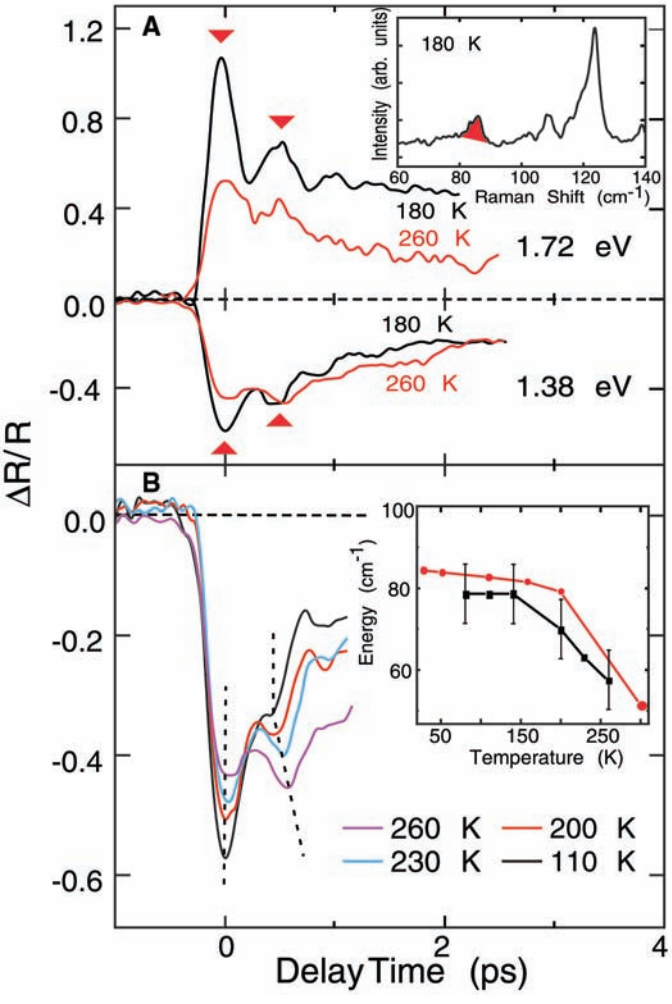


Fig. 3. (**A**) Probe photon energy dependence of the time profile for the $\Delta R/R$ observed at 180 K (black lines) and 260 K (red lines). The probe photon energy was 1.72 and 1.38 eV for the upper and lower panels, respectively. Triangles indicate the peak positions due to the vibratile structure observed at 180 K. (Inset) Raman spectrum in the low-energy region for the I phase observed at 180 K. Temperature dependence of the Raman shift energy for the red-colored mode is plotted in the inset of (**B**). (**B**) Temperature dependence of the time profile for the $\Delta R/R$ observed at 1.38 eV. (Inset) Temperature dependence of the Raman shift (red circles) for the red-colored mode in the inset of (**A**) and that for the vibratile frequency estimated from the time profile of $\Delta R/R$ (black squares). Error bars (mean ± SE) show the observed data fluctuations changing the sample crystal. The red and black lines serve as an eye guide.

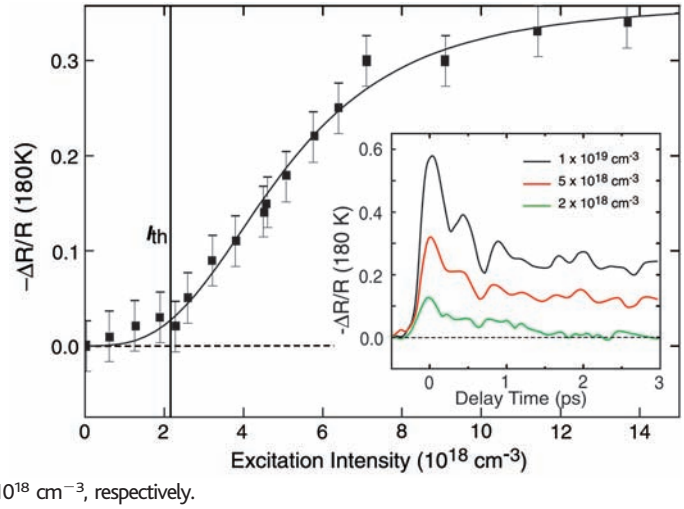


Fig. 4. Excitation intensity dependencies of $\Delta R/R$ observed 3 ps after excitation at 180 K. The black line is an eye guide, and I_{th} indicates the threshold photon density. The fluctuations in $\Delta R/R$ values due to the change of sample crystals even under the same excitation condition are indicated by error bars (mean ± SE) (**Inset**) Time profile of $\Delta R/R$ for various excitation photon densities observed at 180 K. The black, red, and green lines correspond to 1 × 10^{19}, 5 × 10^{18}, and 2 × 10^{18} cm^{-3}, respectively.

$TTF)_2PF_6$ (Fig. 1A). Of course, a further study for mode assignment is necessary to analyze the relation between molecular plane deformation and simple dimer distortion among unit cells, both of which play an important role to drive I-M transition.

One attractive aspect of PIPT is that the photoconversion efficiency shows a nonlinear dependence on the excitation intensity that reflects the role of cooperative interaction (*12*, *13*). In $(EDO\text{-}TTF)_2PF_6$, the photoreflectivity change probed with a 1.38 eV monitor light at 3 ps after photoexcitation showed a threshold-like behavior, with a critical absorbed photon density (I_{th}) of 2×10^{18} cm^{-3} (one photon for every 1500 EDO-TTF molecules) (*19*) (Fig. 4). The estimated heating temperature for the present experimental condition based on the thermodynamic data are only 0.1 K even for a 1×10^{19} cm^{-3} excitation intensity (*24*). Such a highly efficient and nonlinear conversion process is the result of internal cooperative interactions that are similar to other PIPT phenomena (*12*). In addition, the lifetime of the photoinduced M phase strongly depends on the excitation intensity (Fig. 4). This result supports a mechanism in which cooperative interactions occur even in the relaxation process of the photoinduced M phase. In the case of a 2×10^{18} cm^{-3} excitation condition, the $\Delta R/R$ signal disappeared within ~1.5 ps, which appears to facilitate a quick recovery time for applications in a phase-switching device.

To realize a molecular phase-switching device controllable by light with 1 ps (i.e., THz) response time, it is important to develop a material that shows highly sensitive and ultrafast PIPT phenomena at room temperature. One possible way to solve this problem is to use a phase transition due to the cooperation of spin, charge, and orbit degrees of freedom inherent in strongly correlated electrons (*25*–*29*). From our point of view, another way is to use an E-L coupled mechanism mediated by a coherent phonon process with large amplitude (*30*), a possibility discussed in pioneer works of ultrafast photo effect in VO_2 and a neutral-ionic transition system (*20*, *22*, *31*, *32*). In the case of an ordinal photoinduced domino effect, the speed of the domain growth process from photoinjected localized species to meso- or macroscopic regions accompanied by structural changes will be limited by an incoherent phonon process (*13*, *14*). In contrast, if the electronic lattice-coupled change is mediated by an optical phonon mode with k~0 via an E-L interacting channel, the phase change in the meso-size domain mediated by coherent phonons can be induced at once, immediately after photoexcitation.

The present results with $(EDO\text{-}TTF)_2PF_6$ show that this mechanism, the so-called "photo-domino on phonon coherence," seems to be useful for achieving this purpose, even at about room temperature. X-ray structural analysis and soft x-ray emission spectroscopy with femtosecond resolution, in addition to theoretical study, will be needed for clarifying the real mechanism of the observed gigantic photoinduced metallization and also will be important for molecular device–oriented research (*30*, *31*, *33*).

References and Notes

1. D. Jerome, H. J. Schulz, *Adv. Phys.* **51**, 293 (2002).
2. T. Ishiguro, K. Yamaji, G. Saito, Eds., *Organic Superconductors* (Springer-Verlag, Berlin-Heidelberg, ed. 2, 1998).
3. K. Bechgaard, C. S. Jacobsen, K. Mortensen, H. J. Pedersen, N. Thorup, *Solid State Commun.* **33**, 1119 (1980).
4. Y. Shimizu, K. Miyagawa, K. Kanoda, M. Maesato, G. Saito, *Phys. Rev. Lett.* **91**, 107001 (2003).
5. P. Monceau, F. Ya. Nad, S. Brazovskii, *Phys. Rev. Lett.* **86**, 4080 (2001).
6. H. Matsui *et al.*, *J. Phys. Soc. Jpn.* **70**, 2501 (2001).
7. J. E. Hirsch, D. J. Scalapino, *Phys. Rev. Lett.* **50**, 1168 (1983).
8. K. Penc, F. Mila, *Phys. Rev. B* **49**, 9670 (1994).
9. H. Seo, H. Fukuyama, *J. Phys. Soc. Jpn.* **66**, 1249 (1997).
10. D. Schmeltzer, A. Bishop, *Phys. Rev. B* **59**, 4541 (1999).
11. F. O. Karutz, J. U. von Schutz, H. Wachtel, H. C. Wolf, *Phys. Rev. Lett.* **81**, 140 (1998).
12. S. Koshihara, in *Optical Properties of Low-Dimensional Materials*, T. Ogawa, Y. Kanemitsu, Eds. (World Scientific, Singapore, 1998), vol. 2, chap. 3.
13. K. Nasu, Ed., *Photoinduced Phase Transition* (World Scientific, Singapore, 2004).
14. K. Koshino, T. Ogawa, *J. Phys. Soc. Jpn.* **67**, 2174 (1998).
15. A. Ota, H. Yamochi, G. Saito, *J. Mater. Chem.* **12**, 2600 (2002).
16. S. Aoyagi *et al.*, *Angew. Chem.* **43**, 3670 (2004).
17. O. Drozdova, K. Yakushi, A. Ota, H. Yamochi, G. Saito, *Synth. Met.* **133-134**, 277 (2003).
18. O. Drozdova *et al.*, *Phys. Rev. B* **70**, 075107 (2004).
19. Penetration depth at 1.55 eV excitation photon energy was estimated to be approximately 1 μm from the reflection spectrum using the Kramers-Kronig transformation method. The excitation light was assumed to be homogeneously absorbed in the surface region of the crystal with the thickness of 1 μm. As a result, the absorbed photon density used for the observation of photo effect shown in Fig. 2B was estimated to be 6×10^{18} photons/cm^3, corresponding to the excitation intensity of 6.4×10^{14} photons/cm^2. This photon density is approximately equal to a single excitation photon for every 500 donor molecules, based on the unit cell volume of 1170 $Å^3$ at 200 K and including four EDO-TTF molecules.
20. S. Iwai *et al.*, *Phys. Rev. Lett.* **88**, 057402 (2002).
21. T. Ogasawara *et al.*, *Phys. Rev. B* **68**, 180407 (2003).
22. K. Tanimura, I. Akimoto, *J. Lumin.* **94-95**, 483 (2001).
23. The frequency of the Raman mode was determined including the temperature dependence of the Bose factor. A detailed discussion including mode assignment is now in preparation for publication.
24. K. Saito, S. Ikeuchi, A. Ota, H. Yamochi, G. Saito, *Chem. Phys. Lett.*, in press.
25. G. Yu *et al.*, *Phys. Rev. B* **45**, 4964 (1992).
26. M. Fiebig, K. Miyano, Y. Tomioka, Y. Tokura, *Science* **280**, 1925 (1998).
27. T. Ogasawara, K. Tobe, T. Kimura, H. Okamoto, Y. Tokura, *J. Phys. Soc. Jpn.* **71**, 2380 (2002).
28. S. Iwai *et al.*, *Phys. Rev. Lett.* **91**, 057401 (2003).
29. S. Tomimoto, S. Miyasaka, T. Ogasawara, H. Okamoto, Y. Tokura, *Phys. Rev. B* **68**, 035106 (2003).
30. K. Sokolowski-Tintan *et al.*, *Nature* **422**, 287 (2003).
31. A. Cavalleri *et al.*, *Phys. Rev. Lett.* **87**, 237401 (2001).
32. K. Yonemitsu, *J. Phys. Soc. Jpn.* **73**, 2887 (2004).
33. A. Rousse, C. Rischel, J.-C. Gauthier, *Rev. Mod. Phys.* **73**, 17 (2001).

34. The authors are indebted to K. Yakushi and O. Drozdova of the Institute for Molecular Science for their participation in helpful discussions. This work was partially supported by grants-in-aid for scientific research and the 21st Century Center of Excellence programs from the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5706/86/DC1
Movie S1

9 September 2004; accepted 23 November 2004
10.1126/science.1105067

Atom Collision–Induced Resistivity of Carbon Nanotubes

Hugo E. Romero,[1] Kim Bolton,[3] Arne Rosén,[3] Peter C. Eklund[1,2]*

We report the observation of unusually strong and systematic changes in the electron transport in metallic single-walled carbon nanotubes that are undergoing collisions with inert gas atoms or small molecules. At fixed gas temperature and pressure, changes in the resistance and thermopower of thin films are observed that scale as roughly $M^{1/3}$, where M is the mass of the colliding gas species (He, Ar, Ne, Kr, Xe, CH_4, and N_2). Results of molecular dynamics simulations are also presented that show that the maximum deformation of the tube wall upon collision and the total energy transfer between the colliding atom and the nanotube also exhibit a roughly $M^{1/3}$ dependence. It appears that the transient deformation (or dent) in the tube wall may provide a previously unknown scattering mechanism needed to explain the atom collision–induced changes in the electrical transport.

Single-walled carbon nanotubes (SWNTs) with diameters in the range from 1 to 2 nm exhibit quasi–one-dimensional (1D) electronic behavior, including strong van Hove singularities in the electronic density of states (*1*, *2*). Because all the carbon atoms reside at the tube surface, the chemical environment in contact with the nanotube can affect electron transport. The flow of a polar fluid over a nanotube film leads to an induced voltage along the direction of the flow (*3*), and the transconductance of a SWNT field effect transistor (FET)

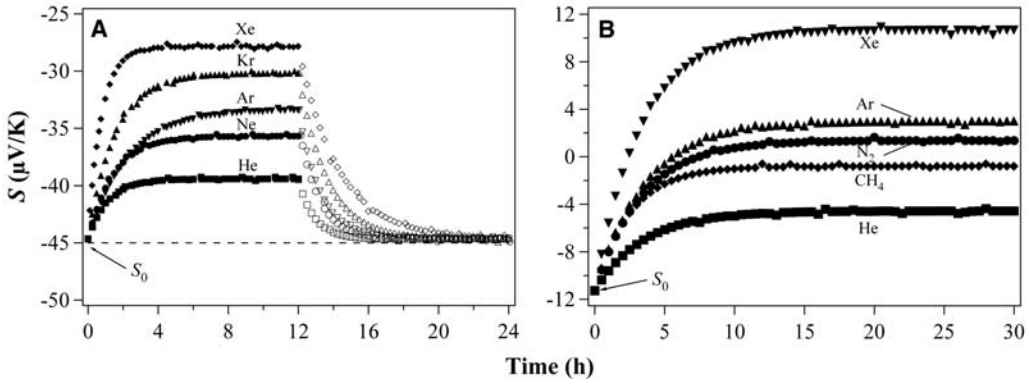


Fig. 1. Time dependence of S at 500 K of (**A**) PLV buckypaper and (**B**) ARC-derived thin film to a sudden overpressure of 1 atm of inert gas (solid symbols). In (**A**), the open symbols show the TEP response after application of a vacuum over the sample.

is sensitive to 2 to 200 parts per million (ppm) NO_2 and NH_3 in contact with the tube wall (*4*). Furthermore, the electrical resistance and thermopower of bundles of SWNTs in films can depend on even the presence of physisorbed ring molecules C_6H_{2x} (where x is an integer from 3 to 6) and are sensitive to the π electron count $(6 - x)$ of the molecule (*5*). Although SWNTs are quite strong in tension, they are easily deformed in the radial direction (*6*). We considered whether the quasi-1D transport can be affected by changes to SWNT lattice itself rather than through bonding or dipolar effects. We have thus undertaken a systematic study of the effect of collisions of inert gas atoms and small molecules (He, Ne, Ar, Kr, Xe, CH_4, and N_2) on the electrical transport properties of SWNTs. Recent electrical transport studies have shown that SWNTs are very sensitive to the presence of He and N_2 (*7*, *8*).

The thermoelectric power (TEP) or Seebeck coefficient, S, and four-probe resistance, R, measurements were carried out on samples in the form of thin films of bundled nanotubes [CarboLex, Incorporated, Lexington, KY; arc-discharge method (ARC)] and purified "buckypaper" [Rice University; pulsed laser vaporization (PLV)] (*9*). The ARC-derived material was also purified (*10*), and thin films were prepared by deposition of a sonicated ethanol solution onto a warm (~50°C) quartz substrate (*11*). The films and buckypaper samples were vacuum annealed at ~1000°C for 12 hours before attaching thermocouples [Chromel (Hoskins Manufacturing Company, Hamburg, MI)-Au/7 atomic % Fe] and electrical (copper) leads with silver epoxy to four corners of the sample for the TEP and resistance measurements. The 2-mm-by-2-mm specimens that contained ropes of 10s to 100s of SWNTs with diameters from 1.0 to 1.6 nm and lengths of several micrometers were placed in a turbo-pumped vacuum chamber ($\sim 1 \times 10^{-7}$ torr), where transport measurements were made in situ in the presence of various gases. The gases were first passed through a purification cartridge (OT-4-SS, R&D Separations, Incorporated, Rancho Cordova, CA) to remove residual O_2 and H_2O. Details of the electrical measurements are available elsewhere (*11*).

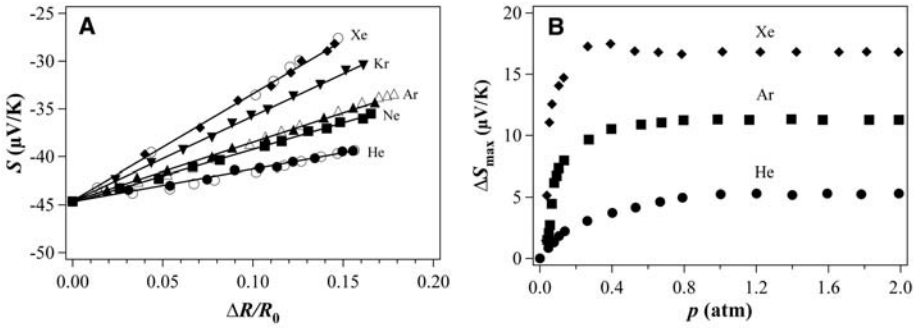


Fig. 2. (**A**) S versus $\Delta R/R_o$ curves showing the effect of inert gases on the transport properties of a SWNT buckypaper prepared from PLV material. The solid symbols are from the time evolution of S and R to 1 atm of gas at 500 K, and the open symbols are from a pressure study at the same temperature, where the maximum responses of S and R to a given pressure were measured. (**B**) Pressure dependence of the maximum change of TEP for the same sample.

The samples were first vacuum-degassed in situ at 500 K to remove adsorbed oxygen and water. As has been seen previously (*8*, *11*, *12*), the TEP decreased slowly over several days from a positive initial value to a negative value representative of the degassed state S_0.

We exposed vacuum-degassed PLV buckypaper (Fig. 1A) and ARC-derived thin film (Fig. 1B) samples to 1 atm of various gases at T = 500 K and measured S and R simultaneously. The samples were degassed in vacuum between successive exposures to the various gases. Under vacuum, S returned slowly to the S_0 value. The differences in S_0 ($S_0 \sim -11$ μV/K for ARC and $S_0 \sim -45$ μV/K for PLV) reflect the different concentrations of tube wall defects and the functionalization introduced during either growth or postsynthesis (acid) purification (*11*, *13*, *14*). The maximum change in thermopower, ΔS_{max}, increased with the mass M of the colliding gas species. The long time constants of the system, which also depend on the mass of the gas atoms or molecules, are consistent with slow diffusion of gas into and out of the pore structure of the SWNT bundles, but the effect of exposure of each gas (1 atm, 500 K) on S is fully reversible.

Previous studies of the TEP behavior of SWNT films are consistent with a diffusion TEP dominated by metallic tubes in a rope (*11*). The TEP of a SWNT rope is modeled as the sum of the conductance-weighted contributions from all of the tubes connected in parallel in the rope (*11*) with the Mott relation (*15*):

$$S = -\frac{\pi^2 k_B^2 T}{3|e|}\left(\frac{d \ln \sigma E}{dE}\right)_{E_F} \quad (1)$$

where σ is the electrical conductivity, e is the electronic charge, k_B is Boltzmann's constant, and T is the temperature. The logarithmic derivative of σ is evaluated at the Fermi energy E_F. As predicted by Eq. 1, S for our purified, annealed, and degassed SWNT samples was observed to be roughly linear in temperature (*16*, *17*). For our purpose here, it is convenient to separate two contributions that limit the conductivity: (i) carrier scattering intrinsic to the degassed metallic tubes (identified with phonons and permanent tube wall defects) and (ii) additional carrier scattering in the metallic tubes associated with the perturbation in the local

[1]Department of Physics, [2]Department of Materials Science and Engineering, Pennsylvania State University, University Park, PA 16802, USA. [3]Department of Experimental Physics, School of Physics and Engineering Physics, Göteborg University and Chalmers University of Technology, SE-412 96, Göteborg, Sweden.

*To whom correspondence should be addressed. E-mail: pce3@psu.edu

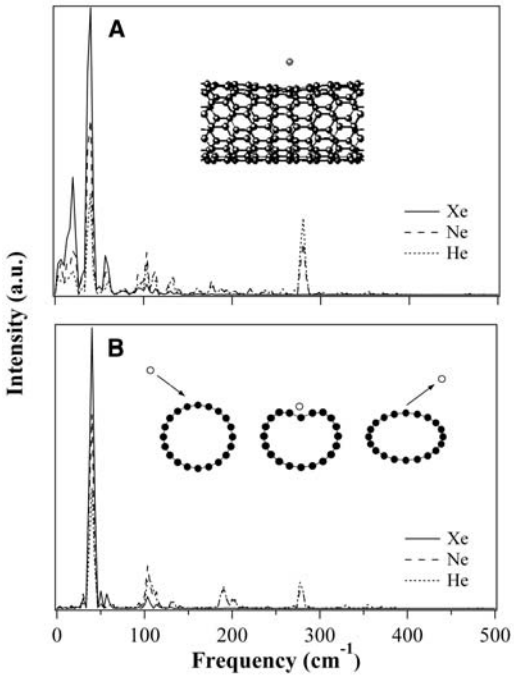


Fig. 3. Computed power spectra of the radial motion of a C atom nearest the point of contact in a (10,0) carbon nanotube at 0 K. The figure shows the phonons induced during (**A**) the first 5 ps of the collision (and includes the gas-tube impact) and (**B**) the second 5 ps after the collision. The inset to (A) shows the side view of a collision between a Xe atom (θ_i = 0, E_i = 13 kcal/mol) and a nanotube. The inset to (B) shows a schematic representation of the tube wall deformation in response to an atom collision. a.u., arbitrary units.

tube wall potential caused by adsorbed gas molecules or collisions with gas molecules. We assume that the scattering contributions (i) and (ii) are additive (*18*):

$$\rho = \rho_0 + \rho_I \quad (2)$$

where ρ_0 is identified with the resistivity of the degassed metallic tubes and ρ_I is identified with the extra impurity scattering due to gas molecules interacting with the tube wall.

Starting with the expression

$$\sigma(E) = e^2 v(E)^2 D(E) \tau(E) \quad (3)$$

where v, D, and τ are, respectively, the free carrier velocity, the density of states, and the carrier lifetime, we find (*7*):

$$S = S_0 - \frac{\pi^2 k_B^2 T}{3|e|}\left(\frac{\rho_I}{\rho_0 + \rho_I}\right) \times \left(\frac{1}{\tau_I}\frac{d\tau_I}{dE} - \frac{1}{\tau_0}\frac{d\tau_0}{dE}\right)_{E_F} \quad (4)$$

Thus, depending on the sign and magnitude of the terms $1/\tau_I(d\tau_I/dE)$ in Eq. 4, we can anticipate a positive or negative slope to the S-versus-ρ_I data collected at fixed temperature T. Furthermore, we can associate $\rho_I/(\rho_0 + \rho_I)$ in Eq. 4 with the fractional change of the four-probe resistance $\Delta R/R_0$, where $\Delta R = R - R_0$ represents the extra resistance caused by the colliding atoms and R_0 is the initial sample resistance in the degassed state. This association further assumes that $R_0 \gg \Delta R$ and that the gas collisions do not change the contact resistance between bundles in the SWNT samples.

We plotted S against $\Delta R/R_0$ for the PLV buckypaper sample exposed to 1 atm of gases at 500 K (Fig. 2A). We found that R exhibited the same time behavior as S and that they are linearly related, i.e., $S \sim \Delta R/R_0$, in agreement with Eq. 4 for constant E_F. The data plotted as S versus $\Delta R/R_0$ (Fig. 2A) show that the slope is related to the mass M of the particular gas. We also found that $S \sim \Delta R/R_0$ (Fig. 2A) for data taken at various pressures in the range $0 < p < 2$ atm; p is a measure of the collision frequency of the atoms with the nanotube walls. Both sets of data fall on the same line for a particular gas, so the slopes of the lines for S versus $\Delta R/R_0$ in Fig. 2A depend on M and not on the chamber pressure.

As shown before, the linear relationship between S and $\Delta R/R_0$ is consistent with the creation of a new scattering channel for the conduction electrons in the metallic tubes, provided that the nanotube E_F remains constant (*7*). In the data presented here, the scattering channel is identified with gas collisions with the tube walls. The variation in slope with M observed in Fig. 2A suggests that the impulse delivered to the tube wall per collision may be an important variable.

We plotted ΔS_{max} versus p for selected inert gases (He, Ar, and Xe) (Fig. 2B). We see that ΔS_{max} saturates with pressure at relatively low pressure (~1 atm) and that the saturation value depends on the gas species (e.g., He, Ar, and Xe). A related saturation of the electrical resistance was also observed (*19*). Resistivity saturation can be associated with many scattering mechanisms and usually occurs when the electron mean free path approaches the interatomic spacing in the material (*20*, *21*). The saturation with pressure seen in Fig. 2B may be caused by a limiting mean free path resulting from increased collisions of the gas atoms with the nanotube wall and might represent the pressure at which neighboring transient deformations in the same tube begin to overlap. With use of a simple kinetic theory for gases, one can derive a relationship between the gas pressure p and the number of colliding atoms (molecules) per surface area unit per unit time:

$$Q = \frac{p}{\sqrt{2\pi m k_B T}} \quad (5)$$

where m is the mass of a single gas atom or molecule. If we designate t as the mean deformation lifetime of a dent and d as the mean diameter of a dent, then the dents will overlap if the area A associated with one collision is $A = \sqrt{3}(d/2)^2$. From Eq. 5, the pressure at which dents overlap and the transport parameters saturate is then given by

$$p_{sat} = \frac{4}{td^2}\sqrt{\frac{2\pi m k_B T}{3}} \quad (6)$$

Molecular dynamics simulations show that the collision-induced phonon excitation persists for as long as ~100 ps and is not sensitive to the mass of the colliding gas for times longer than 10 ps. With use of Eq. 6, we therefore take $t \sim 100$ ps and use the calculated dent diameters $d \sim 15$, 43, and 64 Å to obtain $p_{sat} \sim 0.97$, 0.66, and 0.54 atm for He, Ar, and Xe, respectively, in good agreement with our experimental results of ~0.79, 0.56, and 0.53 atm.

Because of the extreme flexibility of the nanotube wall, we consider what wall deformations might be generated by the collisions of gas atoms or molecules. We also performed molecular dynamics simulations to study this question (*22*) for the case of approaching Xe, Ar, Ne, and He atoms on a finite-length (10,0) carbon nanotube at 0 K (*23*). Calculations were made at 0 K so that complications from the thermal phonon background can be eliminated. Phonons generated in the dent created by a gas collision will propagate away from the collision site and are absorbed in the thermal reservoirs at the tube ends (*24*). To study the transient character of the tube deformations (or phonons) induced by the collision, we followed the displacements of the carbon atoms in a short (400-atom) nanotube over time. We show the power spectrum of the radial C atom motion generated by collisions of Xe, Ne, and He during the first 5 ps of the collision (Fig. 3A), which includes the gas-nanotube

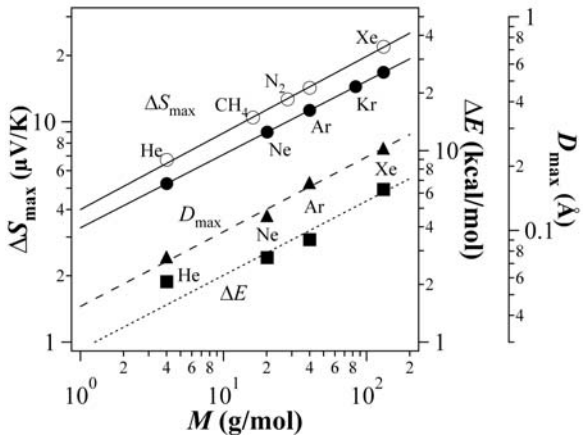


Fig. 4. Maximum TEP change ΔS_{max} of two SWNT samples exposed to gases indicated (ARC, open circles and PLV, solid circles; data from Fig. 1), calculated total energy gained by a (10,0) nanotube upon collision with a gas atom ($\theta_i = 0°$, $E_i = 3.97$ kcal/mol, squares), and maximum radial displacement of the tube C atom immediately after impact with a gas atom ($\theta_i = 45°$, $E_i = 1.99$ kcal/mol, triangles) as a function of the atomic mass of the colliding inert gas. The lines are power-law fits to the data of the forms $\Delta S_{max} = 3.08M^{0.35}$, $\Delta E = 0.91M^{0.39}$, and $\Delta D_{max} = 0.04M^{0.35}$.

impact, and during the second 5 ps (Fig. 3B). The C atom followed was the one nearest to the impact site, but the features of the power spectrum are not sensitive to the C atom chosen. The colliding atoms were incident at $\theta_i = 45°$ to the tube surface normal with an initial energy of $E_i = 13$ kcal/mol. The power spectra (Fig. 3) are weighted so that the area under the spectrum equals the average total energy of the carbon nanotube. The average vibrational energies of the carbon nanotube during the first 5 ps of the collision (Fig. 3A) are 4.2, 3.7, and 2.7 kcal/mol for Xe, Ne, and He, respectively. During the next 5-ps interval (Fig. 3B), we found that the nanotube still retains 2.0, 2.2, and 1.8 kcal/mol for Xe, Ne, and He, respectively.

The most prominent peaks in the power spectra increase in strength with M of the colliding atoms. Snapshots of the tube wall motion show that the gas atom impact locally flattens one side of the tube and that then the entire tube wall near the impact begins to "ring," or oscillate, in an elliptically shaped deformation identifiable with the Raman-active "squash" mode (E_{2g}). The frequency we observe (~43 cm^{-1}) is in reasonable agreement with the calculated squash-mode frequency for a (10,0) nanotube (*25*). The energy dissipation from these squash modes is slow; i.e., the ringing takes place for a long time. These optical modes have zero group velocity, and, for the energy to propagate away from the collision site, the optical squash mode must first decay into two acoustic modes. In all cases, our simulations show that a significant amount of energy remains near the collision site for 10 ps after the collision.

Three features from the power spectra in Fig. 3 are worth mentioning. First, Xe imparts more energy to the carbon nanotube at the time of impact than does Ne, which in turn imparts more energy than He. However, the strong mass dependence we observe for the transferred collisional energy is lost soon after impact. Second, the low-frequency vibrations at ~43 cm^{-1} (squash mode) dominate the vibrational power spectrum, and the energy (amplitude) of this mode is considerably larger after collisions with Xe than after collisions with lighter atoms, e.g., Ne and He. Third, there are fewer peaks in the spectra obtained 5 to 10 ps after collision than in the spectra obtained just after the collider-nanotube impact. Either the energy in the short-lived transient phonons created upon impact flows into the squash-mode phonons, or the energy in these transient phonons flows rapidly along the tube and is absorbed preferentially at the tube ends.

Our simulation studies of collisions between the carbon nanotube and Xe atoms ($\theta_i = 0°$, $E_i = 13$ kcal/mol), suggest that all power spectra (e.g., Fig. 3) should be relatively insensitive to the θ_i of the colliding species. More total energy is transferred at smaller incident angles because a smaller incident angle corresponds to a larger radial component of the gas kinetic energy. Similarly, simulations of collisions between a Xe atom and a carbon nanotube—where $\theta_i = 45°$ and the initial energies were 0.40 kcal/mol (200 K), 0.99 kcal/mol (500 K), and 1.99 kcal/mol (1000 K)—showed that the features of the power spectra are insensitive to the initial kinetic energy. More total energy is transferred to vibrational motion from higher kinetic energy colliders.

The mass dependence of the colliding atoms or molecules on ΔS_{max} is considered in Fig. 4 (1 atm and 500 K; compare with Fig. 1). The straight solid lines in Fig. 4 represent a least squares fit to the experimental ΔS_{max} versus M (circles). Both solid lines exhibit an $\sim M^{1/3}$ dependence. The solid and open circles represent the data for the PLV buckypaper and ARC-derived sample, respectively. For comparison, as observed in our molecular dynamics simulations, we plot the mass dependence of the total energy transfer, ΔE (squares), to a short (10,0) nanotube upon collision (i.e., the total energy lost by the colliding atom), as well as the maximum radial displacement, ΔD_{max} (triangles), of the tube C atom (with which the gas collides) on impact. The slopes of these theoretical lines are insensitive to the simulation conditions (i.e., incident angle, initial energy, and point of impact). The perturbation of the transport properties, the total energy gained by the tube, and the maximum amplitude of the dent obtained from our simulations all share the same approximate power law on the collider mass (i.e., $M^{1/3}$) (*26*). A direct connection between ΔS_{max} and the transient wall deformation requires a microscopic theory for the scattering mechanism. For a static deformation to an elliptical cross section, a metallic tube would become semiconducting if we take the view of recent work on deformations and gap openings at the Fermi level (*27*, *28*). This view would seem to indicate that a local deformation may lead to a relevant direct electronic scattering term that might explain our results. On the other hand, the scattering might proceed via the electron-phonon interaction and the transient, nonthermal local phonon population. Future theoretical work will hopefully address and resolve this issue. Quantitative transport calculations are also needed. A calculation of the change in conductance of a deformed metallic nanotube appears in the work of Lu *et al*. (*27*). This calculation is for a deformation that is uniform along the tube axis and is therefore not a dent. The authors pointed out that tube-tube coupling (such as present in a bundle) very much enhances the effect. Thus, a calculation on an isolated tube may be insufficient.

We propose that the effects we observe in S and R are caused by a new scattering channel for conduction electrons created by collision-induced transient dents in the tube wall. The pressure saturation of the change of the transport parameters (R, S) may be because of the eventual overlap of adjacent dents. The implications of this work for nanotube chemical sensors is simply that gas collisions at a high level can be detected and that they might contribute an important background signal related to the temperature and pressure of the ambient.

References and Notes

1. M. S. Dresselhaus, G. Dresselhaus, P. C. Eklund, *Science of Fullerenes and Carbon Nanotubes* (Academic Press, San Diego, CA, 1996).
2. R. Saito, G. Dresselhaus, M. S. Dresselhaus, *Physical Properties of Carbon Nanotubes* (Imperial College Press, London, 1998).
3. S. Ghosh, A. K. Sood, N. Kumar, *Science* **299**, 1042 (2003); published online 16 January 2003 (10.1126/science.1079080).
4. J. Kong *et al.*, *Science* **287**, 622 (2000).
5. G. U. Sumanasekera *et al.*, *Phys. Rev. Lett.* **89**, 166801 (2002).
6. M.-F. Yu, T. Kowalewski, R. S. Ruoff, *Phys. Rev. Lett.* **85**, 1456 (2000).
7. C. K. W. Adu *et al.*, *Chem. Phys. Lett.* **337**, 31 (2001).
8. G. U. Sumanasekera, C. K. W. Adu, S. Fang, P. C. Eklund, *Phys. Rev. Lett.* **85**, 1096 (2000).
9. A. G. Rinzler *et al.*, *Appl. Phys. A* **67**, 29 (1998).
10. Purification of the ARC-derived material was done in a two-step process: (i) selective oxidation in dry

air for 30 min at 450°C and (ii) refluxing in 4 M HCl at 120°C for 24 hours.
11. H. E. Romero, G. U. Sumanasekera, G. D. Mahan, P. C. Eklund, *Phys. Rev. B* **65**, 205410 (2002).
12. P. G. Collins, K. Bradley, M. Ishigami, A. Zettl, *Science* **287**, 1801 (2000).
13. G. U. Sumanasekera *et al.*, *Phys. Rev. B* **65**, 035408 (2001).
14. We have observed a sample-dependent difference in S_0 that depends on the nature and quality of the purification process. The value depends, for example, on whether HNO_3 or HCl is used to remove residual catalyst from the material. Care must be taken to degas the purified material at ~200°C in a high vacuum (i.e., better than $\sim 10^{-6}$ torr) for 10 to 24 hours and to directly observe the asymptotic approach of S to S_0. In our measurements on about 100 samples mostly derived from electric ARC-produced material, we find that samples with the lowest amorphous carbon content and residual metal catalyst show the smallest, negative S_0 value (i.e., $S_0 \sim -3$ to –8 V/K). The "best" samples had D-band Raman scattering intensity ~1/20 (or less) than that of the T-band scattering, and had measured metal content of ~0.1 to ~0.5 weight % (wt %). Some samples exhibited S_0 as large as –50 V/K. Despite this variation in S_0, all samples that reached an asymptotic negative TEP value to S_0 in the range from –3 to –50 V/K showed essentially the same characteristic response of the transport parameters (R, S) to a specific gas.
15. R. D. Barnard, *Thermoelectricity in Metals and Alloys* (Taylor and Francis, London, 1972).
16. Recently, a broad peak in $S(T)$, observed below 100 K and superimposed on a linear T background, has been attributed to an additional contribution from phonon drag (*11*, *17*). Because our measurements in this study were made at 500 K, we ignore a phonon drag contribution that is a low-temperature effect.
17. J. Vavro *et al.*, *Phys. Rev. Lett.* **90**, 065503 (2003).
18. This assumption is an application of Matthiessen's rule (*15*).
19. H. E. Romero, P. C. Eklund, unpublished data.
20. P. B. Allen, in *Superconductivity in D- and F-Band Metals*, H. Suhl, M. B. Maple, Eds. (Academic Press, New York, 1980).
21. P. L. Rossiter, *The Electrical Resistivity of Metals and Alloys* (Cambridge Univ. Press, Cambridge, 1987).
22. K. Bolton, A. Rosen, *Phys. Chem. Chem. Phys.* **4**, 4481 (2002).
23. The (10,0) tube is semiconducting and has a diameter of 7.83 Å. It is sufficiently small to facilitate the simulation of a large number of scattering events required for statistical analysis.
24. This was done by scaling the velocity of the carbon atoms at the tube ends to zero at each trajectory time step. In this way, energy flowing in the tube axial direction was adsorbed, but not energy that flows in the tube radial or circumferential directions.
25. Saito *et al.* (*2*) reported on the strong diameter dependence of the squash mode. By using their calculated squash-mode frequency for a (10,10) tube, we estimate that the squash-mode frequency for a (10,0) tube is 29 cm^{-1}. In our molecular simulations we observe 43 cm^{-1}. The short length of the nanotube which is clamped at the ends may upshift the squash-mode frequency relative to that obtained for an infinite tube.
26. We also found that the slopes of theoretical lines of the average tube energy and the maximum radial displacement as a function of the colliding mass decrease with time.
27. J.-Q. Lu *et al.*, *Phys. Rev. Lett.* **90**, 156601/1 (2003).
28. C.-J. Park, Y.-H. Kim, K. J. Chang, *Phys. Rev. B* **60**, 10656 (1999).
29. The authors gratefully acknowledge many helpful discussions with G. D. Mahan. This work was supported by NSF (P.C.E.), Pennsylvania State University (P.C.E. and H.R.), and the Swedish Foundation for Strategic Research (A.R. and K.B.).

25 June 2004; accepted 1 December 2004
10.1126/science.1102004

Observation of Large Water-Cluster Anions with Surface-Bound Excess Electrons

J. R. R. Verlet,[1] A. E. Bragg,[1] A. Kammrath,[1] O. Cheshnovsky,[2] D. M. Neumark[1,3]*

Anionic water clusters have long been studied to infer properties of the bulk hydrated electron. We used photoelectron imaging to characterize a class of $(H_2O)_n^-$ and $(D_2O)_n^-$ cluster anions ($n \leq 200$ molecules) with vertical binding energies that are significantly lower than those previously recorded. The data are consistent with a structure in which the excess electron is bound to the surface of the cluster. This result implies that the excess electron in previously observed water-cluster anions, with higher vertical binding energies, was internally solvated. Thus, the properties of those clusters could be extrapolated to those of the bulk hydrated electron.

The hydrated electron (*1*), which is localized and supported within a cavity formed by surrounding water molecules, plays a prominent role in many areas of condensed-phase science such as radiation physics, biological activity, electron transfer, and charge-induced reactivity. Gas-phase clusters of solvent molecules have also been observed to carry an extra charge (*2*), and they represent microscopic analogs to their bulk counterparts, assuming that sufficient solvent molecules are present. For water clusters, however, the critical size beyond which the cluster anion resembles the bulk hydrated electron has remained a controversial issue. Specifically, there are believed to be two localization modes for an excess electron in a water cluster: an internally solvated electron, akin to the bulk; and an electron localized on the surface of the cluster (*3*).

Significant theoretical effort has been devoted to understanding the nature of the excess electron in water clusters, $(H_2O)_n^-$. Path-integral molecular dynamics simulations found that surface states are more stable at sizes up to $n = 32$ molecules, whereas the internal electron is more stable for $n \geq 64$ (*3*). Electron vertical binding energies (VBEs), which correspond to the minimum amount of energy required to remove the electron from the cluster anion with no molecular rearrangement, were predicted to be considerably lower for the surface states than for the internal states. Thus, photoelectron (PE) spectroscopy of $(H_2O)_n^-$ clusters should yield the size at which the surface-to-internal structural transformation occurs.

However, PE spectra measured by Coe *et al.* (*4*) showed no evidence for a surface-to-internal transition in the predicted size range. Instead, the measured VBEs from $(H_2O)_{11}^-$ up to $(H_2O)_{69}^-$ scaled linearly with the inverse of the cluster radius or, equivalently, as $n^{-1/3}$. Such a correlation is expected for an internally solvated electron within a simple dielectric model (*5*), which recovers the observed gradient using the known dielectric constants of bulk water (*4*, *6*). Furthermore, extrapolation of the VBE to infinite size (the bulk) yielded a value of 3.3 eV, which is close to the photoelectric threshold of water, 3.2 eV, estimated by Coe *et al.* (*4*). This result supports the idea that there is an internally solvated electron in these clusters. However, the measured VBEs are in good agreement with VBEs calculated for surface states (*3*), providing the alternative interpretation that surface states were observed in the experiments. There is a similar inconsistency in the electronic absorption spectra of $(H_2O)_n^-$ ($15 \leq n \leq 50$) (*7*), in which the maxima also scale linearly with $n^{-1/3}$ and extrapolate to the observable bulk value (*1*). But, those spectra were also found to be in agreement with the calculated values for surface rather than internal states (*8*). Finally, even though PE spectra of $I^-(H_2O)_n$ clusters showed a VBE-versus-$n^{-1/3}$ correlation with the same gradient as $(H_2O)_n^-$ clusters (*9*), there is consensus that the iodide resides on the surface of small- to medium-sized clusters (*10*, *11*).

The issue of internal versus surface structures has reemerged in the context of two recent time-resolved PE spectroscopy studies focusing on the dynamics of $(H_2O)_n^-$ clusters after electronic excitation (*12*, *13*). Measurement of the internal conversion lifetime of $(H_2O)_n^-$ as a function of size ($n = 25$ to 50) extrapolated to a bulk value of 50 fs, supporting a nonadiabatic

[1]Department of Chemistry, University of California, Berkeley, CA 94720, USA. [2]School of Chemistry, The Sacker Faculty of Exact Sciences, Tel-Aviv University, 69978, Israel. [3]Chemical Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

*To whom correspondence should be addressed. E-mail: dneumark@berkeley.edu

relaxation model for the bulk hydrated electron (*14*). However, extrapolations to the bulk assume that cluster anions are akin to the hydrated electron. Without direct evidence, the question of electron localization in $(H_2O)_n^-$ clusters has remained open.

The nature of electron binding in water-cluster anions can be clarified by locating $(H_2O)_n^-$ isomers with either higher or lower VBEs in the size range studied by Coe *et al.* (*4*). Here, we report observation of a class of water-cluster anions with significantly lower VBEs than those previously reported and assign them to clusters with surface-bound electrons. These clusters were generated by means of a new pulsed valve (*15*) capable of producing very cold molecular beam conditions, and the clusters were characterized using one-photon and time-resolved PE imaging.

Our PE imaging apparatus has been described in detail elsewhere (*16*). Clusters were generated by passing Ar-carrier gas over water (H_2O or D_2O) at 20°C and were injected into the vacuum chamber by means of an Even-Lavie pulsed valve, operated at 100 Hz (*15*). The gas mixture was crossed by high-energy electrons (~500 eV), generating anions through secondary-electron attachment near the throat of the expansion. Anion clusters were then extracted and mass-selected by their time of flight before the laser interaction. Electrons were detached from the cluster by a ~100-fs laser pulse centered at a wavelength of 398 nm ($h\nu$ = 3.11 eV, where h is Planck's constant and ν is the frequency of the laser pulse). The ejected PEs were analyzed by velocity-map imaging (*17*), in which an electrostatic lens projected the three-dimensional (3D) PE velocity distribution onto a focal plane. On this focal plane, a 70-mm multichannel-plate detector coupled to a phosphor screen provided a visual display of the emitted electron cloud, and we captured these images on a charge-coupled device (CCD) camera. PE spectra were then reconstructed from the 2D images using standard methods (*18*).

The one-photon experiments revealed the existence of water-cluster anion isomers with low VBEs. The dynamics of these isomers were characterized by pump-probe experiments, in which an initial pulse of ~100 fs duration and centered at a wavelength of 1650 nm (0.75 eV) pumped the population into an excited state, which was then probed with a delayed probe pulse at a wavelength of 790 nm and of similar time duration.

We acquired PE spectra of a number of $(H_2O)_n^-$ and $(D_2O)_n^-$ clusters, where n = 11 to 200. The effect of varying source backing pressure is shown for $(D_2O)_{50}^-$ in Fig. 1. The black line shows a typical PE spectrum with VBE = 1.78 eV, in agreement with the results of Coe *et al.* (*4*). By gradually increasing the source backing pressure, which generally correlates to cooling of the cluster ensemble, we observed a new feature at lower VBE. At a backing pressure of 70 pounds per square inch (psi), this feature dominates the PE spectrum, and it has a maximum at VBE = 0.97 eV. This observed transition confirms the presence of two distinct species in the molecular beam, labeled as isomers I and II.

A series of PE spectra for $(D_2O)_n^-$ with n = 11 to 150 was collected both at low (30 psi) and high (70 psi) backing pressures (Fig. 2). For clusters with $n > 50$, the PE spectra taken at higher pressure have a single broad asymmetric feature, attributed solely to isomer II. For $n < 50$, isomer I cannot be completely removed in the pressure range studied (30 to 70 psi), and its PE signature increases relative to isomer II as the size is decreased. For these smaller clusters, an additional feature was observed at very small VBE indicated by asterisks in Fig. 2 (isomer III). PE spectra for $(H_2O)_n^-$ are similar to those of $(D_2O)_n^-$ for sizes $n > 50$. Smaller clusters of $(H_2O)_n^-$ do not appear to form isomer II and III as readily, and neither isomer was observed below n = 27.

In Fig. 3, we compare the VBEs of these differing isomers with previously reported VBEs and their $n^{-1/3}$ dependence (*4*, *19*). Our measurements for isomer I are in agreement with those obtained by Coe *et al.* (*4*). Isomer II clusters show consistently smaller VBEs than isomer I clusters, and they scale linearly with $n^{-1/3}$ for $11 \le n \le 20$ and for $n > 50$, with a steeper slope for the set of larger clusters. The VBEs for both isomers II and III appear to connect smoothly, toward smaller sizes, with those previously reported by Kim *et al.* (*19*), who observed two isomers for several $(H_2O)_n^-$ clusters with $n \le 11$. Figure 3 also includes results from molecular dynamics simulations (*3*), illustrating the apparent agreement of the calculated surface VBEs with isomer I.

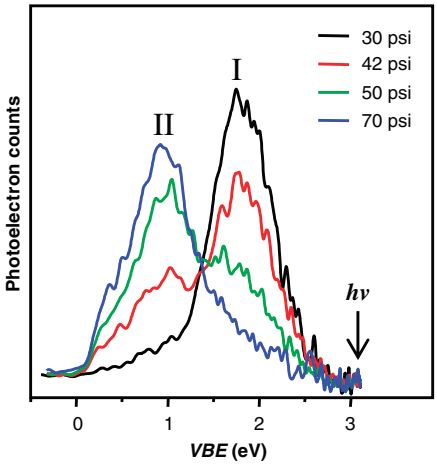


Fig. 1. PE spectra of $(D_2O)_{50}^-$ under varying cluster source conditions. The peaks assigned to internally solvated and surface-localized electron isomers are labeled as I and II, respectively. Different line colors represent PE spectra taken at various backing pressures.

The observation of different isomers over a large range of cluster sizes suggests differing modes of electron localization. Theoretical studies find that the isomer with a lower VBE has its electron localized near the surface of the cluster (*3*), which leads us to assign isomer II to surface states; isomer I is assigned to clusters with internalized electrons. Scaling the VBEs that were calculated from molecular dynamics simulations by 60% for both surface and internal states yielded good agreement with the experimental VBEs for isomers II and I, respectively (Fig. 3). This result implies too strong an electron-water interaction in the model potential used to calculate the VBEs (*20*). The assignment of isomer I to clusters with an internally localized electron agrees with the original assignment by Coe *et al.* (*4*) and validates our extrapolation of the excited-state dynamics in $(H_2O)_n^-$ clusters to those of the bulk hydrated electron (*12*).

Time-resolved pump-probe experiments provided further characterization of isomer II.

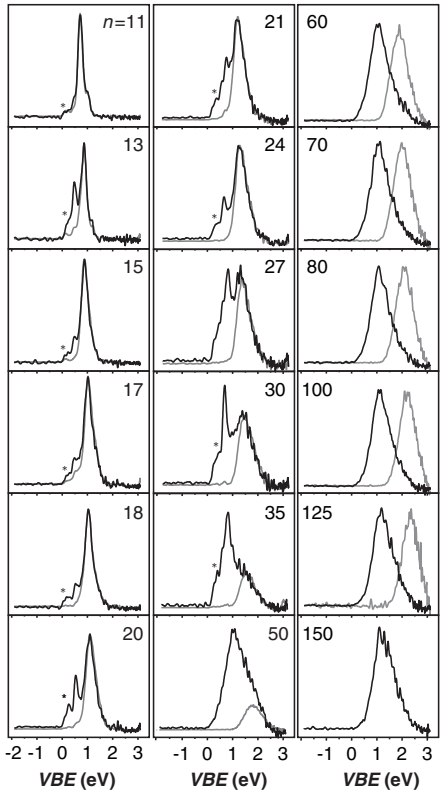


Fig. 2. PE spectra of $(D_2O)_n^-$ with $11 \le n \le 150$ at 30 psi (gray) and 70 psi (black). The peaks marked with asterisks show the presence of a second surface state, isomer III. For n = 11 to 50, the PE feature caused by the internally solvated isomer (gray) is scaled to highlight its contribution to the total PE spectra collected at 70 psi. For n = 60 to 125, there is virtually no contribution from the internally solvated isomer, and the PE spectra for these are scaled to the intensity of the PE spectra from the surface-bound isomer.

In particular, clusters with surface-bound electrons were predicted to support excited states (*8*), and their relaxation rates may thus be quantified by measuring the PE signal at highest electron kinetic energy, corresponding to the time-dependent pump-probe signal (*21*). We measured excited-state lifetimes for $(H_2O)_n^-$ and $(D_2O)_n^-$ clusters with $60 \le n \le 100$ under conditions where only isomer II was present, and compared them to those previously obtained for isomer I (Fig. 4) (*12*). The lifetimes for isomer II were longer and were nearly size-independent. This size invariance suggests a significantly weaker coupling of the electron to the solvent network, compared to the $1/n$ dependence observed for isomer I clusters (*12*). The results in Fig. 4 reinforce our assignments of isomers I and II, because a surface-bound excess electron is expected to be less strongly coupled to the solvent network than an internally localized electron.

The VBEs measured for isomer II of $(D_2O)_n^-$ with $11 \le n \le 20$ merge with those measured by Coe *et al.* (*4*) and Kim *et al.* (*19*) for $(H_2O)_n^-$ with $n \le 11$. This suggests a common structural motif in the two size ranges. Recent theoretical and experimental work (*22*–*24*) indicates that for clusters with $n \le 7$, the lowest energy anion structure is one where a double hydrogen-bond–accepting (AA) water molecule acts as the binding site for the excess electron. The electron is bound by a combination of high dipole moment of the overall solvent network and, more locally, by the two dangling H atoms from the AA water. Our results support a similar electron-binding scheme for isomer II with $11 \le n \le 20$. The evolution of isomer II VBEs for larger cluster sizes, however, suggests a transitional region for $20 < n < 50$, as a new electron-binding motif is established for $n > 50$, characterized by stronger binding and a more pronounced dependence on cluster radius. Only at these large cluster sizes does the surface-bound motif for $(H_2O)_n^-$ clusters become prominent at colder conditions, providing further evidence for a qualitative change in the electron binding once this size is reached. The nature of electron binding to clusters with $n > 50$ is likely to be similar to those observed in simulations (*3*). A local environment involving single acceptor H atoms from a small number of solvent molecules binds the electron on the cluster surface. The electron is further stabilized by the long-range polarization energy induced in the solvent network. The latter effect is demonstrated by the larger gradient in the VBE versus $n^{-1/3}$ plot.

Analogous arguments suggest that isomer III, seen for $(D_2O)_n^-$ with $11 \le n \le 35$, may have similar electron-binding motifs to the lower-VBE isomers observed by Kim *et al.* for $(H_2O)_n^-$ with $n \le 11$ (*19*). By measuring photodetachment cross-sections of the two isomers of $(H_2O)_6^-$, Bailey and Johnson concluded that the electron distribution for the lower-VBE isomer is significantly more diffuse (*25*). This conclusion is supported by ab initio calculations (*23*), predicting a book-like solvent structure for the lower VBE isomer for which the dipole moment of 5.5 D is about half that of the $(H_2O)_6$ structure with the AA water molecule. The evolution of the isomer III VBEs thus suggests a more diffuse surface-electron bound by a number of dangling H atoms. The very weak size dependence of this isomer also suggests that the collective network provides little stabilization because the electron is delocalized over much of the surface.

Regarding the mechanism by which the differing isomers are formed, the considerably higher VBEs for isomer I indicate that it is the more stable anion for $n > 11$. However, isomers II and III are favored by higher source backing pressure, which is generally associated with forming colder and more stable clusters. It appears that isomers II and III are metastable species whose formation is enhanced under colder expansion conditions. These results can be rationalized, given that water-cluster anions are most likely formed in the pulsed jet by the attachment of relatively low-energy electrons to neutral clusters, which then continue to grow as the expansion progresses. As the backing pressure increases, the neutral clusters should be larger and colder in the region of expansion where electron attachment occurs. Hence, under normal conditions similar to those used by Coe *et al.* (*4*) (30 psi in our experiment), we propose that the electrons are attaching to relatively warm, liquid-like clusters, and further solvent molecules may condense onto the cluster after attachment. Furthermore, solvent rearrangement to form the lowest energy anion configuration occurs readily because of the high internal energy available in the solvent network. On the other hand, formation of large anions with surface-localized electrons occurs under much colder conditions, where a small local reorganization can bind the electron on the surface, but the overall low internal energy inhibits the large solvent reorganization necessary to form the internally solvated electron. In that sense, the binding process of the large clusters ($n > 50$) may be viewed as electron attachment to an ice nanocrystal.

For smaller clusters, where $11 \le n \le 20$, the situation is less clear. No neutral $(H_2O)_n$ clusters with AA water molecules have been observed experimentally (*26*). Assuming that the isomer II cluster anions involve an electron binding to an AA water molecule, it appears that these species cannot be formed by simply adding an electron to a neutral cluster; a substantial amount of solvent rearrangement is

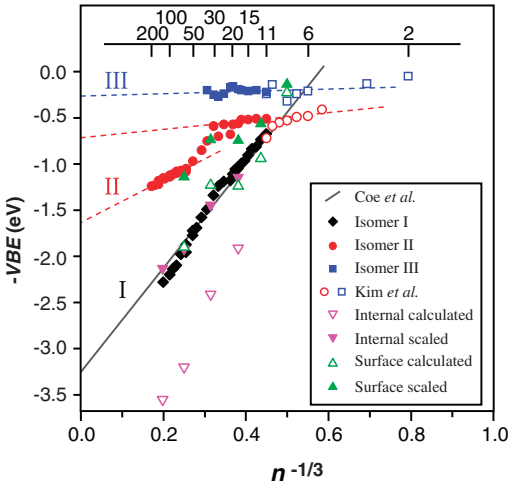


Fig. 3. Plots of VBEs for water-cluster anions versus $n^{-1/3}$. VBEs for isomer I were obtained under conditions similar to those used in previous studies. VBEs for isomers II and III of $(D_2O)_n^-$ were extracted from Fig. 2. Isomer II data also include VBEs obtained for $(H_2O)_n^-$ with $60 \le n \le 200$. The linear fit to data from Coe *et al.* is taken from (*6*). Data from Kim *et al.* (*19*) for clusters with $n \le 11$ are also shown. Data labeled "internal calculated" and "surface calculated" are from molecular dynamics (MD) simulations (*3*); also shown are MD binding energies scaled by 60%.

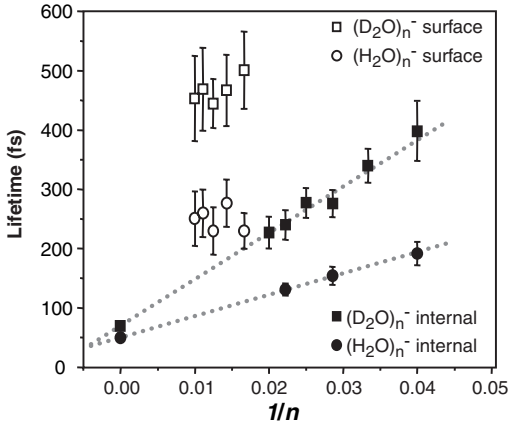


Fig. 4. Pump-probe results for surface and internal states. Relaxation time scales are given after electronic excitation of the surface isomers at 0.75 eV (open symbols) and internal isomers at 1.0 eV (solid symbols) of $(H_2O)_n^-$ (circles) and $(D_2O)_n^-$ (squares). Relaxation dynamics of the internal states follow a $1/n$ size dependence (*12*), whereas dynamics of the surface states are slower and show no size dependence.

required even for a surface state. The observation of isomer II clusters in this size range only for $(D_2O)_n^-$ attests to the fragility of these isomers. They are most likely associated with shallow local minima on the overall potential energy landscape and exist only because of the lower zero-point energies and tunneling rates associated with D_2O as compared to H_2O (*27*). These considerations may explain why, even for larger clusters where we see isomers I and II for both isotopomers, it is easier to make the surface-bound isomer II for $(D_2O)_n^-$ clusters. The preference for isomer I in $(H_2O)_n^-$ clusters could also be a manifestation of the greater stability of the hydrated electron in liquid H_2O than in D_2O (*28*).

References and Notes

1. E. J. Hart, J. W. Boag, *J. Am. Chem. Soc.* **84**, 4090 (1962).
2. H. Haberland, C. Ludewight, H. G. Schindler, D. R. Worsnop, *J. Chem. Phys.* **81**, 3742 (1984).
3. R. N. Barnett, U. Landman, C. L. Cleveland, J. Jortner, *J. Chem. Phys.* **88**, 4429 (1988).
4. J. V. Coe *et al.*, *J. Chem. Phys.* **92**, 3980 (1990).
5. R. N. Barnett, U. Landman, C. L. Cleveland, J. Jortner, *Chem. Phys. Lett.* **145**, 382 (1988).
6. J. V. Coe, *Int. Rev. Phys. Chem.* **20**, 33 (2001).
7. P. Ayotte, M. A. Johnson, *J. Chem. Phys.* **106**, 811 (1997).
8. R. N. Barnett, U. Landman, G. Makov, A. Nitzan, *J. Chem. Phys.* **93**, 6226 (1990).
9. G. Markovich, S. Pollack, R. Giniger, O. Cheshnovsky, *J. Chem. Phys.* **101**, 9344 (1994).
10. W. H. Robertson, M. A. Johnson, *Annu. Rev. Phys. Chem.* **54**, 173 (2003).
11. P. Jungwirth, D. J. Tobias, *J. Phys. Chem. B* **106**, 6361 (2002).
12. A. E. Bragg, J. R. R. Verlet, A. Kammrath, O. Cheshnovsky, D. M. Neumark, *Science* **306**, 669 (2004).
13. D. H. Paik, I.-R. Lee, D. S. Yang, J. S. Baskin, A. H. Zewail, *Science* **306**, 672 (2004).
14. M. S. Pshenichnikov, A. Baltuska, D. A. Wiersma, *Chem. Phys. Lett.* **389**, 171 (2004).
15. U. Even, J. Jortner, D. Noy, N. Lavie, C. Cossart-Magos, *J. Chem. Phys.* **112**, 8068 (2000).
16. A. E. Bragg, J. R. R. Verlet, A. Kammrath, D. M. Neumark, *J. Chem. Phys.* **121**, 3515 (2004).
17. A. T. J. B. Eppink, D. H. Parker, *Rev. Sci. Instrum.* **68**, 3477 (1997).
18. V. Dribinski, A. Ossadtchi, V. A. Mandelshtam, H. Reisler, *Rev. Sci. Instrum.* **73**, 2634 (2002).
19. J. Kim, I. Becker, O. Cheshnovsky, M. A. Johnson, *Chem. Phys. Lett.* **297**, 90 (1998).
20. D. M. Bartels, *J. Chem. Phys.* **115**, 4404 (2001).
21. A. E. Bragg, J. R. R. Verlet, A. Kammrath, O. Cheshnovsky, D. M. Neumark, unpublished data.
22. H. M. Lee, S. B. Suh, K. S. Kim, *J. Chem. Phys.* **118**, 9981 (2003).
23. H. M. Lee, S. Lee, K. S. Kim, *J. Chem. Phys.* **119**, 187 (2003).
24. N. I. Hammer *et al.*, *Science* **306**, 675 (2004).
25. C. G. Bailey, M. A. Johnson, *Chem. Phys. Lett.* **265**, 185 (1997).
26. U. Buck, F. Huisken, *Chem. Rev.* **101**, 205 (2001).
27. F. N. Keutsch, R. J. Saykally, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10533 (2001).
28. P. Han, D. M. Bartels, *J. Phys. Chem.* **95**, 5367 (1991).

29. This research is supported by NSF grant no. CHE-0350585. Additional support from the United States–Israel Binational Science Foundation is gratefully acknowledged.

25 October 2004; accepted 22 November 2004
Published online 16 December 2004;
10.1126/science.1106719
Include this information when citing this paper.

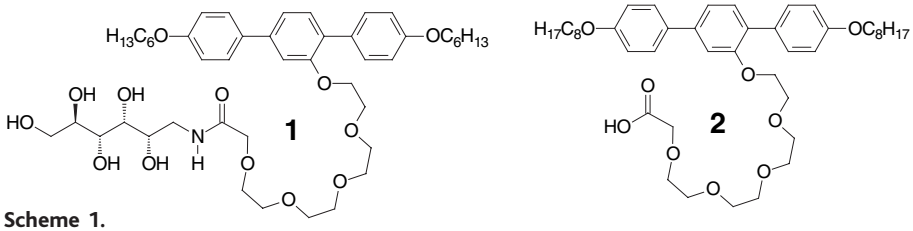
Liquid Crystalline Networks Composed of Pentagonal, Square, and Triangular Cylinders

Bin Chen,[1] Xiangbing Zeng,[3] Ute Baumeister,[2] Goran Ungar,[3] Carsten Tschierske[1]*

T-shaped molecules are designed in such a way that they self-organize into nanoscale liquid crystalline honeycombs based on polygons with any chosen number of sides. One of the phases reported here is a periodic organization of identical pentagonal cylinders; the other one is a structure composed of square-shaped and triangular cylinders in the ratio 2:1. These two different packing motifs represent duals of the same topological class. The generalization of the concept applied here allows the prediction of a whole range of unusual complex liquid crystalline phases.

One of the fascinating aspects of beehive honeycombs is their nearly perfect hexagonal structure. Although it is intuitively "obvious" that this is the way to pack cylinders with the minimum wall area, the mathematical proof came only recently (*1*). Whether we look at living tissue or the organization of nanoparticles, molecules, or atoms, one can find numerous examples of hexagonal patterns. The situation is different for pentagons. In contrast to hexagons, regular pentagons, whose edge lengths and angles are all equal, fail to tile the plane, a fact that has intrigued humans since ancient times (*2*, *3*). However, snugly fitting pentagons can tile the sphere, either by themselves (pentagonal dodecahedron) or in combination with hexagons [e.g., fullerenes (*4*) or supermolecules like polyoxometallates (*5*)]. In contrast, tiling of a flat plane by identical pentagons can be achieved only if regular pentagons are mixed with other polygons in a nonperiodic arrangement [Penrose tilings (*6*, *7*)] or if they contain at least two different angles and side lengths (*8*, *9*). The latter is difficult to realize on a molecular scale with the rather strongly fixed angles and distances dictated by covalent bonding and coordination of standard chemical building blocks. Hence, only a nonplanar two-dimensional (2D) net of pentagons has been realized with a coordination polymer (*10*). We show that the liquid crystalline state (*11*), which combines order and mobility, provides the necessary degree of flexibility for the realization of pentagonal honeycombs.

Liquid crystals (LCs) are known as materials for flat, low power–consuming displays in computing and telecommunication devices. The self-organization within the liquid crystalline phases is driven by the molecular shape, leading to long-range orientational order, and by



Scheme 1.

Table 1. Phase transition temperatures (T) of compounds **1** and **2** (*18*) and the volume fractions of polar side chains (f_P) and nonpolar end chains (f_R). Abbreviations: Cr, crystalline solid state; $Col_{squ}/p4gm$, square columnar phase of the plane group *p4gm*; $Col_{squ}/p4mm$, square columnar phase of the plane group *p4mm*; $Col_{hex}/p6mm$, hexagonal columnar phase of the plane group *p6mm*; Iso, isotropic liquid state.

Compound	T (°C)	f_R	f_P
1	Cr 50 $Col_{squ}/p4gm$ 105 $Col_{hex}/p6mm$ 118 Iso	0.29	0.44
2	Cr 21 $Col_{squ}/p4gm$ 30 $Col_{squ}/p4mm$ 31 Iso	0.41	0.30

[1]Institute of Organic Chemistry, [2]Institute of Physical Chemistry, Martin-Luther-University Halle-Wittenberg, Kurt-Mothes-Strasse 2, D-06120 Halle, Germany. [3]Department of Engineering Materials and Centre for Molecular Materials, University of Sheffield, Sheffield S1 3JD, UK.

*To whom correspondence should be addressed. E-mail: carsten.tschierske@chemie.uni-halle.de

segregation of incompatible parts, leading to the organization of molecules mostly in layers or columns. Usually, the columnar LCs have a column center composed of more-or-less rigid aromatic groups, surrounded by flexible chains. This most often results in a hexagonal 2D periodic organization, and sometimes in a distorted variant, i.e., rectangular or oblique (*12*). Here, in order to create honeycomb-like network structures, the intention is for the rigid rod-like groups to form the outer frame of polygons, with the interior being filled with just enough volume of flexible chains that the perimeter should equal the combined length of a discrete number of rods. The molecules meant to give rise to these LC structures are designed in such a way that the organization favored by their shape is in competition with the drive for microsegregation of their different incompatible parts (amphiphilicity) (*13*–*17*). The T-shaped facial amphiphiles **1** and **2** (*18*) are examples of such molecules. They contain three incompatible segments, a rodlike aromatic core, two terminal lipophilic and flexible alkyl chains, and a polar group in a lateral position.

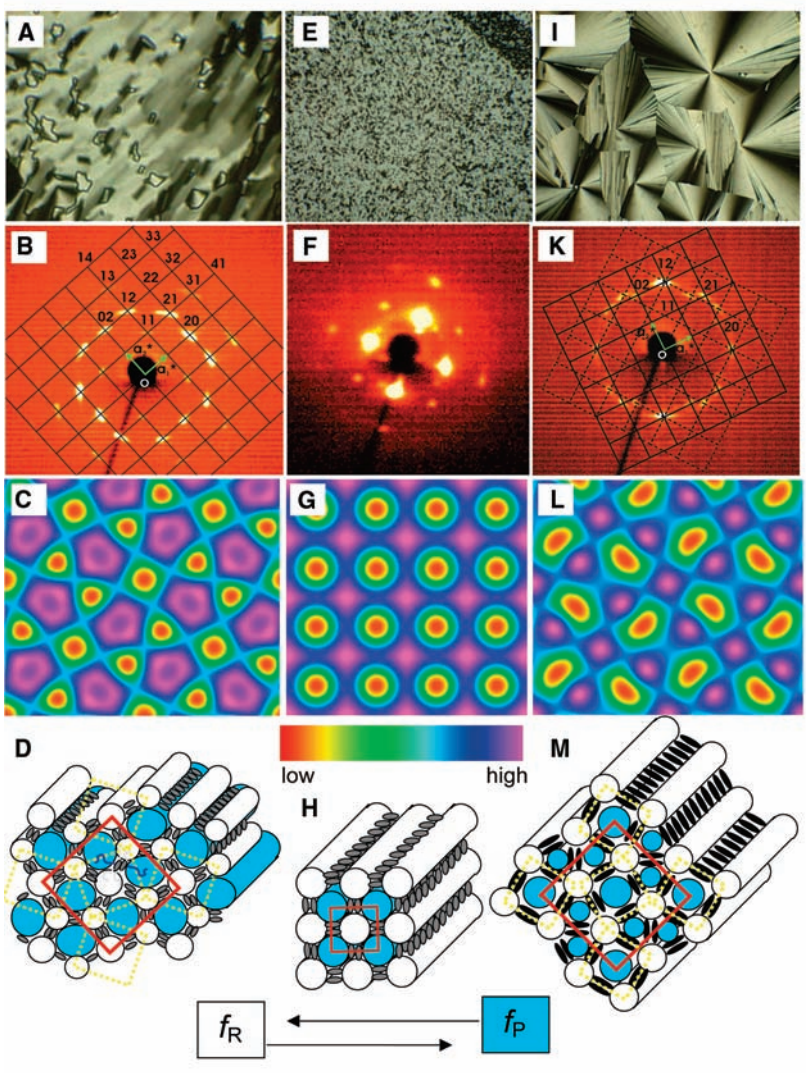


Fig. 1. Mesophases of facial amphiphiles **1** and **2** and their dependence on the volume fractions of the polar (f_P) and nonpolar (f_R) segments. (**A**, **E**, **I**) Typical textures, as seen by microscopy between crossed polarizers. (**B**, **F**, **K**) Typical diffraction patterns of aligned samples (alignment was achieved by slowly cooling a droplet of the sample on a glass substrate from the isotropic liquid state into the liquid crystalline phase; alignment of the LC occurs at the LC-substrate and LC-air interface; the x-ray beam was applied parallel to the substrate). (**C**, **G**, **L**) Reconstructed electron density maps (supporting online text). (**D**, **H**, **M**) Models of molecular organization. (A) Texture of the Col_{squ}/*p4gm* phase of **1** at 80°C; (B) x-ray diffraction pattern of an aligned sample of **1** at 98°C; (C) reconstructed electron density map of the *p4gm* phase of **1**; (D) model of the organization of **1** in the *p4gm* phase; (E) texture of the Col_{squ}/*p4mm* phase of **2** at 30.5°C; (F) typical x-ray diffraction pattern of an aligned sample of the *p4mm* phase (compound related to **1** but with $C_{10}H_{21}$ alkyl chains) (*17*); (G) reconstructed electron density map of the *p4mm* phase of **2**; (H) model of the organization of **2** in the *p4mm* phase; (I) texture of the Col_{squ}/*p4gm* phase of **2** at 29°C; (K) x-ray diffraction pattern of an aligned sample of the *p4gm* phase of **2** at 29°C (two differently oriented domains); (L) reconstructed electron density map of the *p4gm* phase of **2**; (M) model of the organization of **2** in the *p4gm* phase (*22*).

Compound **1**, which has a rather large lateral group, displays two different liquid crystalline phases (Table 1). In addition to a hexagonal columnar (Col_{hex}) high-temperature phase (lattice parameter a_{hex} = 4.4 nm at T = 106°C; figs. S5 and S6 and table S2), a square columnar phase (Col_{squ}) was identified by textural observations (Fig. 1A) and x-ray diffraction from aligned samples. The diffraction pattern of the square columnar phase (Fig. 1B; fig. S4 and table S1) can be indexed on the plane group *p4gm* with a lattice parameter a_{squ} = 8.0 nm at T = 98°C. This Col_{squ}/*p4gm* phase has an unexpectedly large unit cell (*19*) formed by the organization of about 25 or 26 molecules (*20*, *21*). The symmetry of the *p4gm* lattice can be realized by a 90° turn herringbone-like packing of repeat motifs, and the large number of molecules per unit cell suggests that the overall structure is formed by pairs of columns. Pair formation is required only if the individual single columns cannot tile the plane regularly, and hence it was deduced that this mesophase is made up of polar columns, around which the terphenyl units indeed form pentagonal shells. The terminal alkyl chains segregate into columns of their own, located at the corners of the pentagons. These latter columns interconnect the terphenyl units, creating a pentagonal honeycomb network, such that the interior is filled with the flexible polar chains (Fig. 1D) (*22*). This network of pentagons can be described as an array of twin pentagonal cylinders arranged in a herringbone manner (Fig. 2A), leading to the *p4gm* plane group (*23*). This structure was confirmed by electron density calculations based on powder diffraction data obtained with a synchrotron x-ray source. The reconstructed electron density map of the *p4gm* phase of **1** (Fig. 1C) confirms the suggested model (*18*). The high-electron density regions (polar chains, shown in blue and purple) form pentagons that are separated by squares and triangles of low electron density (yellow to red color), which contain the alkyl chains.

From a topological perspective, this tiling of a plane by pentagons can be described as a $(5,^3_4)$ net (*8*). This net has threefold nodes (regions of the alkyl chains interconnecting three cylinder walls), as well as fourfold nodes (regions of the alkyl chains interconnecting four cylinder walls), in the ratio 2:1. The terphenyl units, forming the cylinder walls (*24*), represent the links between the nodes, and the microsegregated regions of the polar chains form pentagonal tiles. This 2D arrangement represents one of the 11 possible Laves tilings (*25*) [Fig. 2, dotted lines and (A)]. Even more interestingly, there is a topological dual of this tiling, in which the positions of the nodes and the tiles are exchanged [Archimedean tiling (*26*); Fig. 2, solid lines and (B)] (*8*). A supra-

molecular organization representing this dual was found for compound **2**. The main difference between compounds **1** and **2** is that the large carbohydrate unit attached to the end of the lateral polyether chain of compound **1** is replaced by a much smaller COOH group in compound **2**. Consequently, the volume fractions of the polar (f_P) and nonpolar (f_R) molecular segments are reversed in **1** and **2**, which changes the mesophase structure.

Compound **2** shows two different liquid crystalline phases (Table 1). The high-temperature phase is a simple square columnar phase with the plane group *p4mm* (Col_{squ}/*p4mm*) and a lattice parameter a_{squ} = 3.5 nm (table S4). The organization within this *p4mm* lattice is shown in Fig. 1, G and H. The polar lateral chains form columns that are framed by four aromatic terphenyl cores. The lipophilic terminal alkyl chains form separate columns interconnecting the aromatic cores end-to-end to yield the overall structure, representing a simple square honeycomb (*22*). The low-temperature phase is a complex columnar phase with a square unit cell, plane group *p4gm* (texture and x-ray diffraction pattern are shown in Fig. 1, I and K) (see also table S3). Its lattice parameter a_{squ} = 7.7 nm is similar to that of the Col_{squ}/*p4gm* phase of **1**. The electron density map of this Col_{squ}/*p4gm* phase indicates the presence of pentagons, squares, and triangles (Fig. 1L). This map can be interpreted as representing either polar columns with a pentagonal cross section, or a combination of polar columns having triangular and square cross sections, depending on the choices of structure factor phases. However, considering that for this compound the volume fraction of the polar chains (f_P = 0.30) is smaller than the volume fraction of the alkyl chains (f_R = 0.41) (Table 1) and that heating (i.e., preferential expansion of the flexible polyether chains in the polar regions) gives the square *p4mm* lattice, it is logical that the polar chains (high electron density) should fill the narrow triangular and square columns, whereas the alkyl chains (low electron density) represent the larger columns with pentagonal shape. Hence, the Col_{squ}/*p4gm* phase of **2** is a liquid crystalline phase composed of triangular and square columns, interconnected by fivefold nodes (Fig. 1M) (*22*). Topologically, this tiling represents one of the 11 Archimedean tilings (*8*), and it is the topological dual of the Laves tiling of pentagons in the *p4gm* phase of **1** (Fig. 2).

The periodic organization of pentagonal cylinders is due to the unique combination of order and mobility in the liquid crystalline fluid state of matter, which allows sufficient deformation of the pentagons as required for this type of regular tiling. In addition to the polyphilic molecular structure, the incorporation of rigid segments with a defined length also seems to be essential for the design of materials capable of this mode of supermolecular organization. The rigid-rod segments tend to restrict the side length of the polygons within relatively narrow limits, giving rise to columns with a well-defined polygonal shape. In contrast to these rigid ternary amphiphiles, flexible ternary amphiphiles, such as ABC star triblock copolymers, can form hexagonal and square columnar arrangements (*27–29*), but not pentagonal ones. Another interesting feature of these mesophase morphologies is their topological duality. The nodes and tiles within the pentagonal tiling of **1** can be exchanged, and this leads to a mesophase structure comprising a regular organization of square and triangular columns for compound **2**. In addition to this topological duality, the exchange of the polar and nonpolar regions leads also to morphological isomerism, designated as color isomers. Hence, there are four possible isomeric mesophase morphologies for the Col_{squ}/*p4gm* phase (Fig. 3, C to F). Color isomers and topological duality are also possible for hexagonal honeycomb organizations (Col_{hex}/*p6mm*). Their topological duals would be triangular columns combined with sixfold nodes (Fig. 3, H and I). For the simple *p4mm* lattice where nodes and tiles have identical symmetry, there are only two color isomers (Fig. 3, A and B). The possible phase structures are summarized in Fig. 3, with only the plane groups of highest symmetry shown. For each of these topologies there can be additional structures with reduced symmetry [e.g., (A) and (B): *p2mm*; (C) to (F): *p2gg*; (G) to (J): *c2mm*]. All these structures can be generalized as edge-to-edge tilings (i.e., nodes are located exclusively at corners of polygons) by identical convex polygons. Thus, there exist 17 types of such tilings by quadrangles, 9 by pentagons, and 13 by hexagons (*8*). Considering the additional possibilities of color-isomers and duals, T-shaped molecules potentially can form a whole range of complex structures (supporting online text and fig. S3).

Pentagonal, square, and trigonal patterns are not restricted to columnar organizations; they can also be found in complex 3D superstructures. For example, the arrangement of polar columns in Fig. 1D (or alkyl columns in Fig. 1M) at the corners of square and triangular tiles in the Col_{squ}/*p4gm* phase is identical to that of spherical dendrimers in the sparsely populated layers (z = 1/4 and 3/4) of the tetragonal 3D liquid crystalline phase $P4_2/mnm$ (*30*). Similarly, the tiling pattern in the Col_{squ}/*p4mm* phase (Fig. 1H), containing only squares, is the same as that of the sparse layers in the $Pm\bar{3}n$ cubic phase in dendrimers. Recently, a liquid quasicrystalline phase was found in spherical supramolecular dendrimers (*31*) having 12-fold rotational symmetry. Its sparse layers are also made up of a mixture of

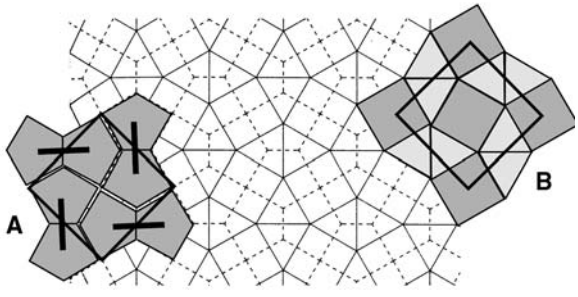


Fig. 2. The dual tilings of the topological class (3^2.4.3.4), representing the plane group *p4gm* (*8*). (**A**) and dashed lines: Laves tiling (*25*) by identical pentagons; (**B**) and solid lines: Archimedean tiling (*26*) by identical squares and triangles.

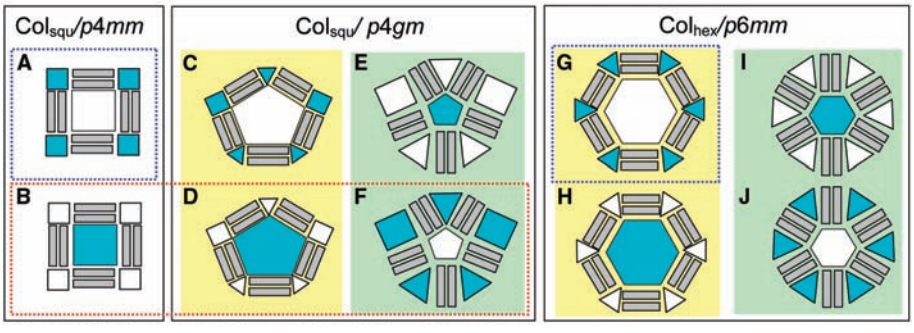


Fig. 3. (**A** to **J**) The topological isomers of the tiling of the plane in the different honeycomb-like network structures formed by different types of T-shaped molecules, facial amphiphiles (lower row), and rodlike molecules with two polar end groups and lateral nonpolar chains (bolaamphiphiles, upper row). Blue and white areas represent polar and nonpolar regions, respectively, whereas the rodlike cores are shown in gray. The arrangements in the lower row represent the color-isomers of the corresponding arrangements in the upper row. Topological duals are indicated by green and yellow backgrounds, respectively. Cross-relations between the duals [e.g., (C) and (F)] are characterized by a 90° rotation of the rodlike cores. The morphologies framed by a red dotted line are reported here, those framed by blue dotted lines have been found for bolaamphiphiles (*14, 15*), whereas the nonframed morphologies have yet to be discovered.

squares and triangles, but these are tiled in a quasi-periodic rather than truly periodic manner. This analogy raises the intriguing possibility that an equivalent quasiperiodic columnar phase might be discovered in the future.

References and Notes

1. T. C. Hales, *Discrete Comput. Geom.* **25**, 1 (2001).
2. J. Simon, P. Bassoul, *Design of Molecular Materials: Supramolecular Engineering* (Wiley, Chinchester, UK, 2000).
3. D. L. D. Caspar, E. Fontano, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14271 (1996).
4. M. Deza, P. W. Fowler, A. Rassat, K. M. Rogers, *J. Chem. Inf. Comput. Sci.* **40**, 550 (2000).
5. A. Müller *et al.*, *Angew. Chem. Int. Ed. Engl.* **37**, 3359 (1999).
6. R. Penrose, *Bull. Inst. Math. Appl.* **10**, 266 (1974).
7. R. Penrose, *Eureka* **39**, 16 (1978).
8. B. Grünbaum, G. C. Shephard, *Tilings and Patterns* (Freeman, New York, 1987).
9. T. Sugimoto, T. Ogawa, *Forma* **15**, 75 (2000).
10. B. Moulton, J. Lu, M. J. Zaworotko, *J. Am. Chem. Soc.* **123**, 9224 (2001).
11. P. J. Collings, M. Hird, *Introduction to Liquid Crystals: Chemistry and Physics* (Taylor and Francis, London, 1997).
12. R. J. Bushhy, O. R. Lozman, *Curr. Opin. Colloid Interface Sci.* **4**, 343 (2002).
13. C. Tschierske, *J. Mater. Chem.* **11**, 2647 (2002).
14. M. Kölbel *et al.*, *J. Am. Chem. Soc.* **123**, 6809 (2001).
15. X. H. Cheng *et al.*, *J. Am. Chem. Soc.* **125**, 10977 (2003).
16. B. Chen *et al.*, *Angew. Chem. Int. Ed. Engl.* **43**, 4540 (2004).
17. B. Chen *et al.*, *J. Am. Chem. Soc.* **116**, 8608 (2004).
18. Materials and methods (synthesis, analytical data, and details of x-ray scattering) are available as supporting material on *Science* Online.
19. The lattice parameter is about 2.7 times the molecular length of L = 3.0 nm, as measured using molecular models (CPK models) for the most extended conformation between the two ends of the terminal alkyl chains.
20. From the lattice parameter it was calculated according to $n = V_{mol}/ha_{squ}^{2}$ that about 25.5 molecules are arranged in the unit cell with a height of h = 0.45 nm. The height h was obtained from the maximum of the diffuse wide-angle scattering, which corresponds to the mean lateral distance between adjacent molecules. For the calculation of the molecular volumina (V_{mol}), the crystal volume increments of Immirzi (*21*) were used.
21. A. Immirzi, B. Perini, *Acta Crystallogr. Sect. A* **33**, 216 (1977).
22. The mesophases reported here represent ordered fluids, which implies a high degree of conformational, rotational, and translational mobility. Hence, there are no well-defined crystal-like structures, as might be suggested by the models shown. Instead, in these highly dynamic systems, there are only distinct regions with enhanced concentrations of aromatic cores, regions of the polar hydrogen bonding groups, and regions of the nonpolar segments. Depending on the degree of incompatibility between the segments, the interfaces between these segregated regions can be quite sharp or more diffuse. Also, because of this nanoscale disorder, the number of molecules within a distinct volume region is only an average value.
23. This structure corresponds to the pentagonal isohedral tiling *P5-24* (*8*).
24. About two terphenyl units on average are arranged in the diameter of the cylinder walls.
25. There are 11 monohedral tilings of a plane with regular vertices. A vertex is called regular if the angle between each consecutive pair of edges is $2\pi/v$, where v is the number of edges meeting at a vertex of a tiling. These 11 tilings are often called Laves tilings after the crystallographer Fritz Laves (*8*).
26. There are precisely 11 distinct edge-to-edge tilings of a plane by regular polygons such that all vertices are of the same type. These tilings are called Archimedean tilings (*8*).
27. S. Sioula, N. Hadjichristidis, E. L. Thomas, *Macromolecules* **31**, 8429 (1998).
28. H. Hückstädt, A. Göpfert, V. Abetz, *Macromol. Chem. Phys.* **201**, 296 (2000).
29. K. Yamauchi *et al.*, *Macromolecules* **36**, 6962 (2003).
30. G. Ungar, Y. Liu, X. B. Zeng, V. Percec, W.-D. Cho, *Science* **299**, 1208 (2003).
31. X. B. Zeng *et al.*, *Nature* **428**, 157 (2004).

32. This work was supported by the Deutsche Forschungsgemeinschaft (GRK 894), the Fonds der Chemischen Industrie, and the European Community (RTN LCDD). We thank A. Gleeson for helping to set up the experiment at Daresbury Synchrotron and CCLRC (Council for the Central Laboratory of the Research Councils) for granting the beamtime.

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5706/96/DC1
Materials and Methods
SOM Text
Scheme S1
Figs. S1 to S6
Tables S1 to S4
References

23 September 2004; accepted 29 November 2004

Electron Tunneling Through Organic Molecules in Frozen Glasses

Oliver S. Wenger, Brian S. Leigh, Randy M. Villahermosa, Harry B. Gray,* Jay R. Winkler*

Reaction rates extracted from measurements of donor luminescence quenching by randomly dispersed electron acceptors reveal an exponential decay constant of 1.23 per angstrom for electron tunneling through a frozen toluene glass (with a barrier to tunneling of 1.4 electron volts). The decay constant is 1.62 per angstrom (the barrier, 2.6 electron volts) in a frozen 2-methyl-tetrahydrofuran glass. Comparison to decay constants for tunneling across covalently linked xylyl (0.76 per angstrom) and alkyl (1.0 per angstrom) bridges leads to the conclusion that tunneling between solvent molecules separated by ~2 angstroms (van der Waals contact) is 20 to 50 times slower than tunneling through a comparable length of a covalently bonded bridge. Our results provide experimental confirmation that covalently bonded pathways can facilitate electron flow through folded polypeptide structures.

Electron tunneling processes have been found to play pivotal roles in solid-state physics (*1*), chemistry (*2*), and biology (*3*, *4*). In molecular systems, electron tunneling is commonly described in terms of superexchange theory (*5*, *6*). The efficiency of electronic coupling across a tunneling barrier is captured by the exponential decay constant β, which is determined experimentally through investigations of the distance dependence of electron transfer (ET) rates. Superexchange theory in the two-state approximation suggests that β depends on the size of each bridge unit (δ), the strength of the coupling between them (h_{bb}), and the difference in energy (Δε) between the tunneling electron and the reduced or oxidized states of the bridge (*7*).

$$\beta = \left(\frac{2}{\delta}\right)\ln\left(\frac{\Delta\varepsilon}{h_{bb}}\right) = \left(1.025\text{Å}^{-1}\text{eV}^{-1/2}\right)\sqrt{\Delta E_{eff}} \quad (1)$$

It is difficult to attribute differences in decay constants to specific properties of the bridge, because superexchange parameters are often difficult to measure directly, although they can be estimated. Instead, we can characterize electron tunneling systems in terms of effective barrier heights ΔE_{eff} (Eq. 1).

Most investigations of ET have focused on two redox centers that are covalently attached to variable-length bridges, and the resulting β values are usually interpreted in terms of the electronic properties of the bridge. Saturated hydrocarbon spacers typically exhibit β ~ 1.0 $Å^{-1}$ (where ΔE_{eff} = 0.95 eV, and 1 eV is ~23 kcal/mol) (*8*–*10*). The decay constant for phenylene bridges depends on the dihedral angle between adjacent aromatic rings [0.4 to 0.8 $Å^{-1}$ (*11*, *12*); ΔE_{eff} = 0.15 to 0.61 eV]. Polyene and phenylenevinylene bridges exhibit efficient ET over very long distances: β values as low as 0.04 $Å^{-1}$ (ΔE_{eff} = 0.002 eV) have been reported (*13*). In the latter systems, bridge-state energies and hence barrier heights vary with changes in the length of the bridge, and electron transport (hopping rather than tunneling) is likely to be involved. It is curious, given the tunneling energy dependence expressed in Eq. 1, that β values appear to be primarily functions of the bridge composition and are relatively independent of the properties of the redox centers (*14*).

Electron tunneling bridges need not be constructed of purely covalent linkages. Extensive investigations of long-range ET in proteins provide compelling support for the notion that electronic coupling pathways

Beckman Institute, California Institute of Technology, Pasadena, CA 91125, USA.

*To whom correspondence should be addressed. E-mail: hbgray@caltech.edu (H.B.G.) and winklerj@caltech.edu (J.R.W.)

involve both covalent links as well as van der Waals and hydrogen-bonded contacts between strands of the polypeptide (*4*, *15*–*17*). Such studies suggest that electrons will take a longer path through bonds and avoid shorter paths that involve nonbonded contacts. However, direct studies of ET through nonbonded contacts that would provide appropriate β values have been lacking. Electron tunneling through vacuum or through solvent molecules represents a limiting case: There are no covalent links between redox sites and the bridge, and no covalent links between bridge elements. In vacuum, β values are estimated to range from 3.0 to 4.0 $Å^{-1}$ (*18*). Studies of ET in C-shaped donor-acceptor complexes (*19*), ultrafast ET quenching in fluid solution (*20*, *21*), and luminescence quenching in frozen solvents (*22*–*24*) demonstrate that covalent bridges are not required for long-range electron tunneling. Experiments in fluid solvents provide information about relatively short tunneling distances, because translational diffusion of electron donors **D** and electron acceptors **A** is faster than long-range tunneling. Prior reports have suggested that decay constants for tunneling through frozen solvents [e.g., 2-methyl-tetrahydrofuran (MTHF), β = 1.2 $Å^{-1}$ (*25*), and glycerol, β = 1.4 $Å^{-1}$ (*26*)] are slightly greater than those found for analogous covalent bridges.

Studies in frozen solvents are particularly amenable to systematic investigations of the electronic and structural factors that determine ΔE_{eff} values. In vitrified solvents at liquid-nitrogen temperature, excited-state ET reactions must be highly optimized in order to compete with other radiative and nonradiative decay processes. We have found that $[Ir(\mu\text{-pyrazolyl})(1,5\text{-cyclooctadiene})]_2$ {$[Ir(\mu\text{-pz})(COD)]_2$ or $\mathbf{D_{Ir}}$} and 2,6-dichloro-1,4-benzoquinone ($\mathbf{A_Q}$) are well-suited for ET in frozen solvents (*27*, *28*). The lifetime (τ_o) of the iridium dimer $^3B_2$ excited state donor ($^*\mathbf{D_{Ir}}$) is 3.2 μs in MTHF and toluene glasses at 77 K; this relatively long lifetime has allowed us to probe ET over distances as great as 20 Å. A driving force of ~1.6 eV in fluid acetonitrile solution (*29*–*31*) compensates for the retarding effects of low temperature and solvent rigidity on the $^*\mathbf{D_{Ir}} \rightarrow \mathbf{A_Q}$ ET reaction. In cryogenic matrices, it is likely that $^*\mathbf{D_{Ir}} \rightarrow \mathbf{A_Q}$ ET involves both nuclear and electronic tunneling in the inverted driving force regime where reactions exhibit only a modest dependence on the standard free energy change $\Delta G°$ (*32*). Moderate concentrations of $\mathbf{A_Q}$ (0.05 to 0.20 M) produce significant reductions in luminescence quantum yields as well as accelerated, highly nonexponential $^*\mathbf{D_{Ir}}$ decays (Fig. 1). This quenching is attributed to electron transfer from $^*\mathbf{D_{Ir}}$ to $\mathbf{A_Q}$ (*33*, *34*). Inspection of the data clearly reveals that electron tunneling through toluene (Fig. 1C) is more efficient than tunneling through MTHF (Fig. 1B).

Electron tunneling parameters for the $^*\mathbf{D_{Ir}} \rightarrow \mathbf{A_Q}$ ET reaction have been obtained from analysis of the $^*\mathbf{D_{Ir}}$ decay kinetics. If $\mathbf{D_{Ir}}$ and $\mathbf{A_Q}$ are randomly distributed in the glasses, as is expected for charge-neutral molecules, and if ET rates vary exponentially with **D**-**A** distance (r), an expression that depends on two adjustable parameters can be used to fit the luminescence decay data [$I(t)$]: the distance decay constant β and the ET rate (k_{ET}^{o}) when **D** and **A** are in contact ($r = r_o$) (*35*–*37*).

$$\ln\left(\frac{I(t)}{I(t=0)}\right) = -\frac{t}{\tau_o} - \frac{[A]}{(132.12\ M\ Å^3)} \times \int_{r_o}^{\infty}\{1 - \exp[-k_{ET}^{o}\, t\, \exp[-\beta(r - r_o)]]\}r^2\, dr \quad (2)$$

where the constant 132.12 is defined for **A** concentrations M. In order to decouple these two parameters, it is necessary properly to scale the amplitudes of the luminescence decay data using relative luminescence quantum yields measured at 77 K (*20*, *38*). Because of uncertainties in the measurements of quantum yields (±10%), we could not obtain adequate fits with a single pair of β and k_{ET}^{o} values for each solvent. Nevertheless, the values for the two parameters span relatively small ranges, and the results confirm the qualitative conclusion that toluene mediates long-range electron tunneling more efficiently than MTHF (Fig. 1) (*39*).

The $^*\mathbf{D}$ decay kinetics can be transformed to represent the effective tunneling distance (R_{eff}) as a function of time (*25*). The tunneling distance should be independent of quencher concentration and, if the tunneling rates are exponential functions of the **D**-**A** distance (r), plots of R_{eff} as a function of $\ln(t)$ should be linear with slopes of 1/β.

$$R_{eff}(t) = -\left(\frac{3}{4\pi[A]}\right)^{\frac{1}{3}}\left[\ln\left(\frac{I(t)}{I(t=0)}\right) + \frac{t}{\tau_o}\right]^{\frac{1}{3}} = r_o + \frac{1}{\beta}\ln(g\, k_{ET}^{o}) + \frac{1}{\beta}\ln(t) \quad (3)$$

where g is equal to 1.9. When transformed according to Eq. 3, the $^*\mathbf{D_{Ir}}$ tunneling-distance plots are indeed linear, but are not precisely superimposed, presumably because of uncertainties in the low temperature quantum yield measurements used to scale the data (Fig. 2A). If we introduce adjustable scaling factors (constrained to lie within the ±10% scaling-error boundaries), we can nicely overlay the tunneling distance data and obtain global sets of β and k_{ET}^{o} values for each solvent (toluene, β = 1.23 $Å^{-1}$, $k_{ET}^{o} = 5.0 \times 10^{12}\ s^{-1}$; MTHF, β = 1.62 $Å^{-1}$, $k_{ET}^{o} = 4.6 \times 10^{13}\ s^{-1}$) (Fig. 2B). We also have performed this analysis on tunneling data collected with Ru(2,2′:6,2″-terpyridine)$_2^{2+}$ ($\mathbf{D_{Ru}}$) and $Fe(OH_2)_6^{3+}$ ($\mathbf{A_{Fe}}$) in aqueous H_2SO_4 (25% v/v) glasses (β = 1.59 $Å^{-1}$, $k_{ET}^{o} = 8.5 \times 10^{12}\ s^{-1}$) (*38*, *40*).

A coupling limited rate/distance plot (Fig. 3) reveals that tunneling 20 Å through vitreous toluene is ~450 times faster than tunneling through an aqueous glass (25% H_2SO_4) and more than 750 times faster than tunneling through an MTHF glass. Tunneling across comparable lengths of covalently linked bridges tends to be much more efficient. Exponential distance decay constants for tunneling across alkane bridges are on the order of 1.0 $Å^{-1}$ (*10*). We have found that the distance decay constant for tunneling from a *para*-methoxy-*N*,*N*-dimethylaniline (DMA)

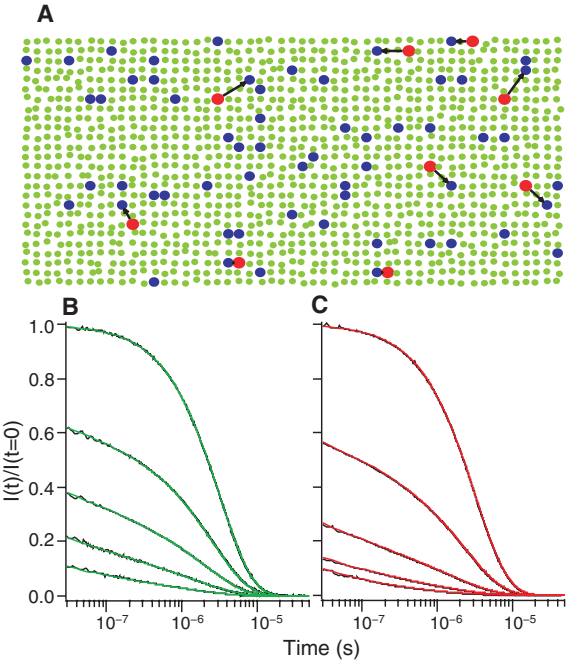


Fig. 1. (**A**) ET from photoexcited donors (red) to randomly dispersed acceptors (blue) in a rigid solvent (green) glass will produce nonexponential excited-state decays, because of the wide distribution of **D**-**A** distances present in the sample. (**B** and **C**) Luminescence decay kinetics (black) for $[Ir(\mu\text{-pz})(COD)]_2$ (3×10^{-5} M) in (B) MTHF and (C) toluene glasses at 77 K in the presence of 2,6-dichloro-1,4-benzoquinone (upper to lower traces, $[\mathbf{A_Q}]$ = 0.0, 0.05, 0.10, 0.15, and 0.20 M). The smooth lines (green and red) are decays calculated with Eq. 2 (*35*–*37*), with r_o = 7 Å and the following parameters ($[\mathbf{A_Q}]$, β, k_{ET}^{o}): MTHF, 0.05 M, 1.50 $Å^{-1}$, $1.2 \times 10^{13}\ s^{-1}$; 0.10 M, 1.62 $Å^{-1}$, $4.1 \times 10^{13}\ s^{-1}$; 0.15 M, 1.64 $Å^{-1}$, $8.3 \times 10^{13}\ s^{-1}$; 0.20 M, 1.65 $Å^{-1}$, $18.0 \times 10^{13}\ s^{-1}$. Toluene, 0.05 M, 1.23 $Å^{-1}$, $0.3 \times 10^{13}\ s^{-1}$; 0.10 M, 1.32 $Å^{-1}$, $2.5 \times 10^{13}\ s^{-1}$; 0.15 M, 1.33 $Å^{-1}$, $2.7 \times 10^{13}\ s^{-1}$; 0.20 M, 1.19 $Å^{-1}$, $0.3 \times 10^{13}\ s^{-1}$.

donor to a Ru(2,2′-bipyridine)$_3^{3+}$ [Ru(bpy)$_3^{3+}$] acceptor linked by oligoxylene bridges is 0.76 Å^{-1} (*12*, *41*, *42*). Again taking 20 Å as a reference distance, we find that tunneling across an alkane bridge is almost 40,000 times faster than tunneling through vitreous MTHF. In the case of aromatic bridges, 20-Å tunneling across xylenes is almost 3000 times faster than tunneling through a toluene glass.

The ΔE_{eff} = 1.4 eV for electron tunneling from $^*\mathbf{D}_{Ir}$ to $\mathbf{A}_Q$ through toluene is substantially lower than the 2.6-eV barrier in MTHF. The barrier height for $^*\mathbf{D}_{Ru}$-$\mathbf{A}_{Fe}$ tunneling in aqueous sulfuric acid glasses (ΔE_{eff} = 2.4 eV) is only slightly lower than the MTHF barrier. Superexchange theory suggests that the differences in barrier heights can be attributed to variations in tunneling energy gaps ($\Delta\varepsilon$), coupling between solvent molecules (h_{bb}), and the size of the bridge repeat unit (δ). Because $\Delta\varepsilon$ is a vertical energy gap at the ET transition state configuration, it cannot readily be extracted by experiment. The energies of the $^*\mathbf{D}^{+/0}$ and $\mathbf{A}^{0/-}$ states can be estimated from electrode potentials, corrected for the reduced dielectric constant in the solvent glass. The energies of the oxidized and reduced states of the solvents can be estimated from gas-phase ionization potentials and electron affinities, adjusted for the reduced solvent glass dielectric response. There are substantial uncertainties in an analysis of this type, but for all three solvents, the tunneling energy gaps are estimated to be at least 3 eV. Moreover, the energy gaps for tunneling mediated by reduced states of the intervening solvent molecules are somewhat smaller than for the oxidized states. The more efficient tunneling in toluene glasses can be attributed to the presence of π-bonding and antibonding orbitals that lead to a smaller tunneling energy gap than that in MTHF or H_2O/H_2SO_4.

Two additional factors are likely to play a role in determining the relative tunneling distance-decay factors in these frozen solvents: δ and h_{bb}. An excess electron or hole in toluene will be delocalized over the set of π molecular orbitals, but in MTHF and H_2O, a charge is likely to be localized primarily on two or three atoms. Hence, the observed trend in β values is broadly consistent with the size of the bridge units. In toluene and MTHF, coupling between bridge units is mediated by van der Waals contacts, whereas the aqueous glass is interlaced with strong hydrogen bonds. Hydrogen bonds between solvent molecules in the aqueous glass may compensate for a large energy gap to produce a tunneling barrier on par with that of MTHF.

The distance-decay constants for tunneling through MTHF and toluene speak to the efficiency of coupling across van der Waals contacts. Electrochemical measurements of tunneling across self-assembled alkanethiol monolayers have produced β values of ~1.0 Å^{-1} (*10*); each additional carbon atom decreases the tunneling rate by a factor of 2.4. Electron tunneling across a single MTHF molecule should have a decay factor similar to that of an alkane bridge. The 1.62 Å^{-1} distance decay factor for the MTHF glass suggests that there is a substantial coupling penalty associated with tunneling across the contacts between solvent molecules. Radial distribution functions extracted from molecular dynamics simulations of THF suggest that the intermolecular contact distance is ~2 Å (*43*). If we approximate the MTHF glass as a face-centered cubic lattice of spherical particles, then the nearest-neighbor center-to-center distance will be 6.185 Å. If the tunneling across each 4.185-Å diameter sphere has the same distance decay as tunneling along an alkane chain (β ~1.0 Å^{-1}), then we can infer that tunneling across an MTHF molecule decreases the rate by a factor of 65, and tunneling across the 2-Å van der Waals gap between molecules decreases the tunneling rate by a factor of 340. A similar analysis for toluene suggests that the rates drop by a factor of 26 across the 4.297-Å molecule, and by a factor of 90 across the 2-Å gap (*44*). The tunneling decays across these van der Waals contacts translate into exponential distance decay constants of 2.9 to 4.0 Å^{-1}. As might be expected, these values are in the range suggested for tunneling through a vacuum (*18*).

Our findings have profound implications, particularly for electron tunneling processes in proteins. Electron tunneling along the polypeptide backbone in a β-strand exhibits an exponential distance dependence with a decay constant (1.1 Å^{-1}) (*16*) close to that found for tunneling through alkane chains (1.0 Å^{-1}) (*8*–*10*). In general, however, electron tunneling rates in proteins do not scale monotonically with distance (Fig. 3). Rates at a single **D**-**A** separation can differ by as much as a factor of 10^3 and **D**-**A** distances that differ by as much as 5 Å can produce identical rates (*4*). These variations are the direct results of the structure of the poly-

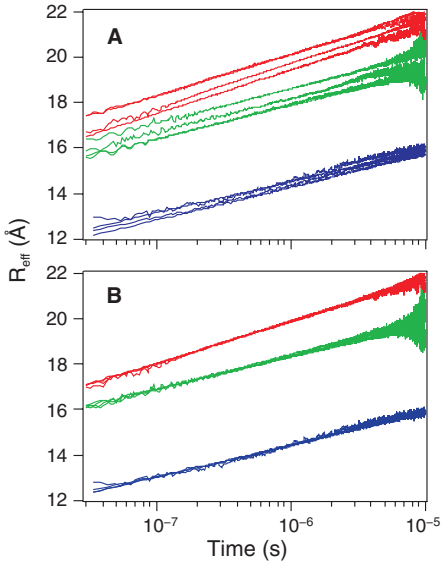


Fig. 2. (**A**) Plot of the effective tunneling length (R_{eff}, Eq. 3) as a function of time after laser excitation for ET from $^*\mathbf{D}_{Ir}$ to $\mathbf{A}_Q$ in toluene (red) and MTHF (green) glasses and ET from $^*\mathbf{D}_{Ru}$ to $\mathbf{A}_{Fe}$ in a 25% (v/v) aqueous H_2SO_4 glass (blue). (**B**) Plot of R_{eff} as a function of time with data scaled to optimize overlap of the different concentration curves. Acceptor concentrations: [$\mathbf{A}_Q$], 0.05, 0.10, 0.15, and 0.20 M; [$\mathbf{A}_{Fe}$], 0.05, 0.10, 0.25, and 0.50 M. β and k_{ET}° values: $^*\mathbf{D}_{Ir}$-$\mathbf{A}_Q$, MTHF, r_o = 7 Å, 1.62 Å^{-1}, 4.6 × 10^{13} s^{-1}; $^*\mathbf{D}_{Ir}$-$\mathbf{A}_Q$, toluene, r_o = 7 Å, 1.23 Å^{-1}, 5.0 × 10^{12} s^{-1}; $^*\mathbf{D}_{Ru}$-$\mathbf{A}_{Fe}$, 25% H_2SO_4, r_o = 4 Å, 1.59 Å^{-1}, 8.5 × 10^{12} s^{-1}.

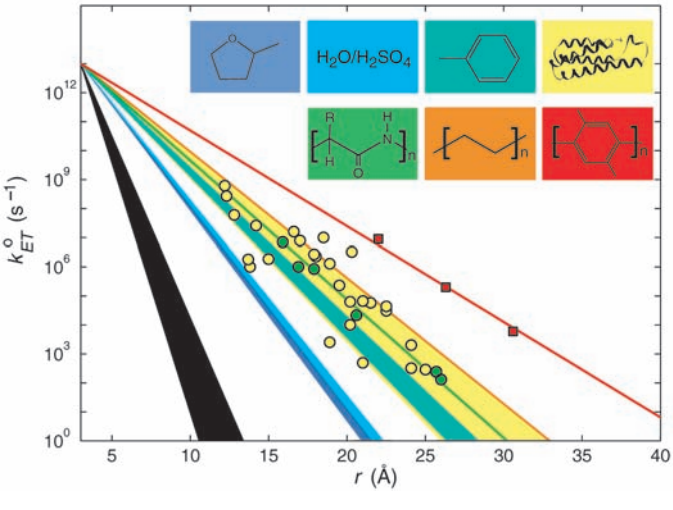


Fig. 3. Timetable for activationless electron tunneling through various media: vacuum (black, β = 2.9 to 4.0 Å^{-1}) (*18*), MTHF glass (dark blue, β = 1.57 to 1.67 Å^{-1}), aqueous glass (light blue, β = 1.55 to 1.65 Å^{-1}) (*38*), and toluene glass (teal, β = 1.18 to 1.28 Å^{-1}). Investigations of ET rates in **D**-(bridge)-**A** complexes have produced exponential distance dependences: xylyl bridges, β = 0.76 Å^{-1} (red squares) (*12*); alkane bridges, β = 1.0 Å^{-1} (orange) (*10*); and β-strand bridges in ruthenium-modified azurin, β = 1.1 Å^{-1} (green circles) (*4*). Tunneling rates for a variety of ruthenium-modified proteins (cytochrome c, cytochrome b_{562}, myoglobin, and high-potential iron-sulfur protein) (*4*) and for Zn-cytochrome c/Fe-cytochrome c crystals (*46*) are indicated by yellow circles in the yellow wedge. Insets show formulas or structures of molecules in the intervening media.

peptide medium separating **D** and **A** (*45*). Poor coupling across nonbonded interfaces will produce a strong bias in favor of covalent and hydrogen-bonded pathways between redox sites (*4*, *17*, *18*). The coupling disparity between bonded and nonbonded interfaces accounts in large part for the finding that protein ET rates do not exhibit a uniform dependence on distance, but instead depend critically on the composition of the medium between redox sites (*4*, *17*). Nearly all of the 32 protein ET rates shown in Fig. 3 are bounded by the MTHF glass and covalent alkane bridge decays. This behavior is consistent with tunneling pathway considerations: The best coupling routes in proteins resemble those in alkane bridges; the poorest pathways will have many gaps, as in an MTHF glass. Clearly, the protein fold is the key determinant of biological electron transfer rates; it establishes not only the distance between redox centers, but also the balance between the bonded and nonbonded contacts that mediate electronic coupling.

References and Notes

1. L. Esaki, *Science* **183**, 1149 (1974).
2. J. R. Miller, *J. Phys. Chem.* **79**, 1070 (1975).
3. D. De Vault, J. H. Parkes, B. Chance, *Nature* **215**, 642 (1967).
4. H. B. Gray, J. R. Winkler, *Q. Rev. Biophys.* **36**, 341 (2003).
5. H. M. McConnell, *J. Chem. Phys.* **35**, 508 (1961).
6. J. Halpern, L. E. Orgel, *Disc. Faraday Soc.* **1960**, 32 (1960).
7. S. S. Skourtis, D. N. Beratan, *Adv. Chem. Phys.* **106**, 377 (1999).
8. H. Oevering *et al.*, *J. Am. Chem. Soc.* **109**, 3258 (1987).
9. M. D. Johnson, J. R. Miller, N. S. Green, G. L. Closs, *J. Phys. Chem.* **93**, 1173 (1989).
10. J. F. Smalley *et al.*, *J. Am. Chem. Soc.* **125**, 2004 (2003).
11. A. Helms, D. Heiler, G. McLendon, *J. Am. Chem. Soc.* **114**, 6227 (1992).
12. R. Villahermosa, thesis, California Institute of Technology (2002).
13. W. B. Davis, W. A. Svec, M. A. Ratner, M. R. Wasielewski, *Nature* **396**, 60 (1998).
14. F. D. Lewis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12536 (2002).
15. D. S. Wuttke, M. J. Bjerrum, J. R. Winkler, H. B. Gray, *Science* **256**, 1007 (1992).
16. R. Langen *et al.*, *Science* **268**, 1733 (1995).
17. H. B. Gray, J. R. Winkler, *Annu. Rev. Biochem.* **65**, 537 (1996).
18. J. N. Onuchic, D. N. Beratan, J. R. Winkler, H. B. Gray, *Annu. Rev. Biophys. Biomol. Struct.* **21**, 349 (1992).
19. K. Kumar, I. V. Kurnikov, D. N. Beratan, D. H. Waldeck, M. B. Zimmt, *J. Phys. Chem. A* **102**, 5529 (1998).
20. K. Weidemaier, H. L. Tavernier, S. F. Swallen, M. D. Fayer, *J. Phys. Chem. A* **101**, 1887 (1997).
21. L. Burel, M. Mostafavi, S. Murata, M. Tachiya, *J. Phys. Chem. A* **103**, 5882 (1999).
22. J. R. Miller, J. A. Peeples, M. J. Schmitt, G. L. Closs, *J. Am. Chem. Soc.* **104**, 6488 (1982).
23. R. C. Dorfman, Y. Lin, M. D. Fayer, *J. Phys. Chem.* **93**, 6388 (1989).
24. T. Guarr, M. E. McGuire, G. McLendon, *J. Am. Chem. Soc.* **107**, 5104 (1985).
25. J. R. Miller, J. V. Beitz, R. K. Huddleston, *J. Am. Chem. Soc.* **106**, 5057 (1984).
26. S. Strauch, G. McLendon, M. McGuire, T. Guarr, *J. Phys. Chem.* **87**, 3579 (1983).
27. Preparation of the Ir(I) complex followed a published procedure (*28*). 2,6-Dichloro-1,4-benzoquinone (Sigma-Aldrich) was recrystallized from methanol before use. All solvents were high-purity grade; they were dried and distilled before use.
28. J. L. Atwood *et al.*, *Inorg. Chem.* **23**, 4050 (1984).
29. Estimated from published redox potentials of $[Ir(\mu\text{-pz})(COD)]_2$ (*30*) and 2,6-dichloro-1,4-benzoquinone (*31*).
30. D. C. Smith, H. B. Gray, *Coord. Chem. Rev.* **100**, 169 (1990).
31. S. Fukuzumi, S. Koumitsu, K. Hironaka, T. Tanaka, *J. Am. Chem. Soc.* **109**, 305 (1987).
32. B. S. Brunschwig, N. Sutin, *Comments Inorg. Chem.* **6**, 209 (1987).
33. Multipole-multipole energy transfer quenching (*34*) of the $^3B_2$ excited state of the iridium dimer will be negligible, because there is no spectral overlap between $[Ir(\mu\text{-pz})(COD)]_2$ $^3B_2$ luminescence (>600 nm) and 2,6-dichloro-1,4-benzoquinone absorption (<520 nm).
34. T. Förster, *Ann. Phys. (Leipzig)* **2**, 55 (1948).
35. M. Inokuti, F. Hirayama, *J. Chem. Phys.* **43**, 1978 (1965).
36. A. Blumen, J. Manz, *J. Chem. Phys.* **71**, 4694 (1979).
37. A. Blumen, *J. Chem. Phys.* **72**, 2632 (1980).
38. A. Ponce, H. B. Gray, J. R. Winkler, *J. Am. Chem. Soc.* **122**, 8187 (2000).
39. A. M. Napper, H. Liu, D. H. Waldeck, *J. Phys. Chem. B* **105**, 7699 (2001).
40. The effective tunneling distance is a **D-A** center-to-center measure; consequently, the larger $*D_{Ir}$-A_Q pair exhibits longer tunneling distances than $*D_{Ru}$-A_{Fe}.
41. ET across xylyl bridges was measured with a flash-quench technique (*42*). $Ru(bpy)_3^{3+}$-$(xylyl)_n$-DMA (where n = 3 to 5) was generated in acetonitrile by quenching laser-excited $Ru(bpy)_3^{2+}$-$(xylyl)_n$-DMA with methylviologen. Subsequent DMA to $Ru(bpy)_3^{3+}$ ET was monitored by transient absorption spectroscopy (*12*). ET rates varied with the number of xyyl groups in the bridge: $9.0 \pm 0.3 \times 10^6$ s^{-1}, n = 3; $2 \pm 1 \times 10^5$ s^{-1}, n = 4; $6 \pm 1 \times 10^3$ s^{-1}, n = 5.
42. I.-J. Chang, H. B. Gray, J. R. Winkler, *J. Am. Chem. Soc.* **113**, 7056 (1991).
43. W. Drabowicz, *Z. Naturforsch.* **45**, 1342 (1990).
44. W. L. Jorgensen, E. R. Laird, T. B. Nguyen, J. Tirado-Rives, *J. Comp. Chem.* **14**, 206 (1993).
45. I. A. Balabin, J. N. Onuchic, *Science* **290**, 114 (2000).
46. F. A. Tezcan, B. R. Crane, J. R. Winkler, H. B. Gray, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5002 (2001).

47. We thank J. Kim, J. Lee, and J. Magyar for several helpful discussions. Supported by BP, the NSF (grant no. CHE-0078809), and the Arnold and Mabel Beckman Foundation. O.S.W. acknowledges a postdoctoral fellowship from the Swiss National Science Foundation and B.S.L., a graduate fellowship from the Parsons Foundation.

9 August 2004; accepted 23 November 2004
10.1126/science.1103818

Coral ^{230}Th Dating of the Imposition of a Ritual Control Hierarchy in Precontact Hawaii

Patrick V. Kirch[1]* and Warren D. Sharp[2]

In proto-historic Hawaii (1500–1795 A.D.), as in many other evolving polities, temples functioned as centers for control over production and the extraction of surplus food and goods. Thorium-230 dates (uncertainty ± ~10 years) on branch coral dedicatory offerings from temples in the Kahikinui district (Maui) indicate that its temple system was constructed within 60 years, far more rapidly than indicated by radiocarbon dating. Introduction of the temple system in 1580–1640 A.D. coincided with predatory expansion and consolidation of the Maui polity to form an incipient archaic state.

Complex chiefdoms and the early archaic states that emerged from them typically exercised power through the imposition of a control hierarchy with three to four administrative levels. Linked to religious ideology and temple ritual (*1*, *2*), these hierarchies are reflected archaeologically in monumental architecture, particularly temples and palaces (*3*). The Hawaiian Islands offer a model locality for investigating the transition from a chiefdom to an archaic state because of their isolation from outside contact after ~1200–1400 A.D. (*4*, *5*). During the late expansion to proto-historic periods of Hawaiian prehistory (~1500–1800 A.D.), several sociopolitical changes took place, including (i) the emergence of endogamous classes and of the ideology of kingship, (ii) the replacement of kinship-based land control with a territorial system, (iii) the imposition of corvée labor and the collection of surplus food and goods as tribute, (iv) agricultural and aquacultural intensification, and (v) the imposition of ritualized controls on production (*6*–*9*). The rate at which these changes took place has been uncertain.

Hawaiian temples were constructed on monumental-scale platforms and terraces of dry-laid basalt. The temple system corresponded to a hierarchy of major gods associated with agricultural production or war (*10*). Agricultural temples (ranging in size from 50 to 2000 m^2) often marked territorial land divisions. War temples (up to ~10,000 m^2) are more dispersed (*6*).

Dating the construction of temples has depended on radiocarbon (^{14}C) dating of wood charcoal. Dated Maui Island temples (*11*, *12*) increased in size between ~1400 and 1650 A.D., corresponding to a period of archipelago-wide population growth (*13*) and the intensification of irrigated and dryland agricultural systems (*14*). ^{14}C dating in the past 500 years, however, is confounded by irregular fluctuations in atmospheric ^{14}C concentrations, leading to uncertainties of 40 to

[1]Department of Anthropology, 232 Kroeber Hall, University of California, Berkeley, CA 94720, USA. [2]Berkeley Geochronology Center, Berkeley, CA 94709, USA.

*To whom correspondence should be addressed. E-mail: kirch@sscl.berkeley.edu

>250 years (*15*, *16*). Moreover, spuriously old ^{14}C ages can result from burning of wood that significantly predates temple building.

We report a refined chronology of construction for temples on Maui and Molokai Islands obtained via ^{238}U-^{234}U-^{230}Th dating (^{230}Th dating) of branch corals used as dedicatory offerings. This approach offers substantial advantages over ^{14}C dating. ^{230}Th ages of materials that remain a closed system and are free of initial ^{230}Th depend only on their $^{230}Th/^{238}U$ and $^{234}U/^{238}U$ ratios and the decay constants of ^{230}Th, ^{234}U, and ^{238}U: The isotope ratios may be accurately determined by mass spectrometry, and the decay constants are known to ~0.3% or better (*17*). Corrections for initial ^{230}Th in corals are generally minor.

The ancient Hawaiians regarded branch coral (several species in the genus *Pocillopora*) as an appropriate offering material for placement on coastal fishing shrines and at inland agricultural temples (Fig. 1, A and B). The preservation of delicate surface structures (verrucae) on these corals and the lack of evidence of abrasion or erosion indicate that the corals were collected as live specimens before their placement at temples (Fig. 1C); the measured age of the sample therefore corresponds to the target event to be dated: the placement of a dedicatory offering (*18*, *19*). That corals were not simply deposited at temples after construction is indicated by their incorporation into walls and platform fill.

We focused on temples within the ancient district of Kahikinui, on southeast Maui Island, where 30 temple foundations are preserved, with basal foundations ranging from 60 to 1400 m^2 in area (*20*). Most were agricultural temples, located in areas of productive soils for dryland agriculture (*21*). We selected branch coral offerings from seven of these temples, representing different temple size classes and both coastal and inland temples (Table 1). We also dated corals from a boundary temple at Kawela, Molokai Island, which was at various times incorporated into the competing Maui and Oahu polities. The Kawela site is interpreted as a *Hale o Lono*, or a temple dedicated to the deity Lono and linked to the annual Makahiki rite of tribute collection, one of the means of extracting agricultural surplus that was used by the chiefly classes (*22*). We also analyzed one living coral to refine our estimate of initial ^{230}Th (*23*). The time span of prehistoric Hawaiian occupation in Kahikinui has been established as 1400–1800 A.D. through ^{14}C dating of 159 samples of wood charcoal from agricultural, habitation, and ceremonial contexts (supporting online text and fig. S1).

To determine whether temple construction and dedication could be more accurately dated, nine corals from archaeological sites (Table 1) and one living coral from Maui were analyzed for U and Th isotopes. The archaeological corals consist of broken branches, basal fragments of coral colonies, and an intact coral colony (Table 2). Many samples have distinct upper and lower surfaces, apparently reflecting long exposure in a constant orientation to dust infall, rain, and sunlight. The samples are identified as *Pocillopora meandrina*, except for one of the Molokai corals, which is probably *P. damicornis*. *P. meandrina* is a common coral species on Hawaiian reef slopes at shallow depths and grows at a rate of ~1 cm/year. Individual colonies can be as large as 30 to 40 cm (*24*).

When shallow-water corals form, they are profoundly depleted in ^{230}Th relative to its ultimate parent, ^{238}U; therefore, they are generally highly suitable for ^{230}Th dating (*25*). The ^{232}Th concentrations of corals analyzed for this study are, however, about an order of magnitude higher than ^{232}Th concentrations reported for many reef-building corals (*25*). The presence of ^{232}Th indicates that small amounts of nonradiogenic ^{230}Th ($^{230}Th_{nr}$) (^{230}Th not generated by in situ U decay) are also present. Such $^{230}Th_{nr}$ can affect calculated ages for corals that are less than a few thousand years old (*26*). The effect of $^{230}Th_{nr}$ can be corrected by using ^{232}Th (half-life $t_{1/2}$ ~ 14 Gy) as an index isotope if the $^{230}Th_{nr}/^{232}Th$ ratio can be estimated or determined from a sample of known age. Sources of $^{230}Th_{nr}$ in the Hawaiian corals might include silicate detritus derived from basalt or eolian dust (*27*) and Th from seawater that is either dissolved or adsorbed on suspended particles. Expected $^{230}Th_{nr}/^{232}Th$ atomic ratios in such materials range from $\sim 4 \times 10^{-6}$ in silicates (assuming secular equilibrium and a typical crustal Th/U ratio) to $\sim 1 \times 10^{-5}$ in shallow Pacific seawater (*28*). Other sources of Th and combinations of sources are also possible.

To estimate the $^{230}Th_{nr}/^{232}Th$ ratio, we analyzed a sample from the outer centimeter of a living coral collected adjacent to the Maui study area. Given typical *P. meandrina* growth rates of ~1 cm/year, this sample is expected to be about 0.5 years old. Adopting that age (and subtracting the ^{230}Th formed between collection and analysis), we calculate for the modern coral a $^{230}Th_{nr}/^{232}Th$ atomic ratio of $1.4 \pm 0.2 \times 10^{-5}$, or equivalently a $^{230}Th_{nr}/^{232}Th$ activity ratio of 2.5 ± 0.4. This value falls near the middle of the range of $^{230}Th_{nr}/^{232}Th$ values found in an extensive study of living and young fossil corals in the central Pacific (*26*).

The ^{232}Th concentration of the modern coral sample (336 pg/g) is similar to or greater than the ^{232}Th concentrations of the other dated Maui corals; thus, there is no evidence of ^{232}Th addition to the fossil corals during their subaerial exposure. Accordingly, we used the $^{230}Th_{nr}/^{232}Th$ ratio determined from the modern coral to correct for $^{230}Th_{nr}$ in the fossil corals. The largest $^{230}Th_{nr}$ correction applied (to sample Kawela 1-A) was 18 years; the median correction for all fossils corals was

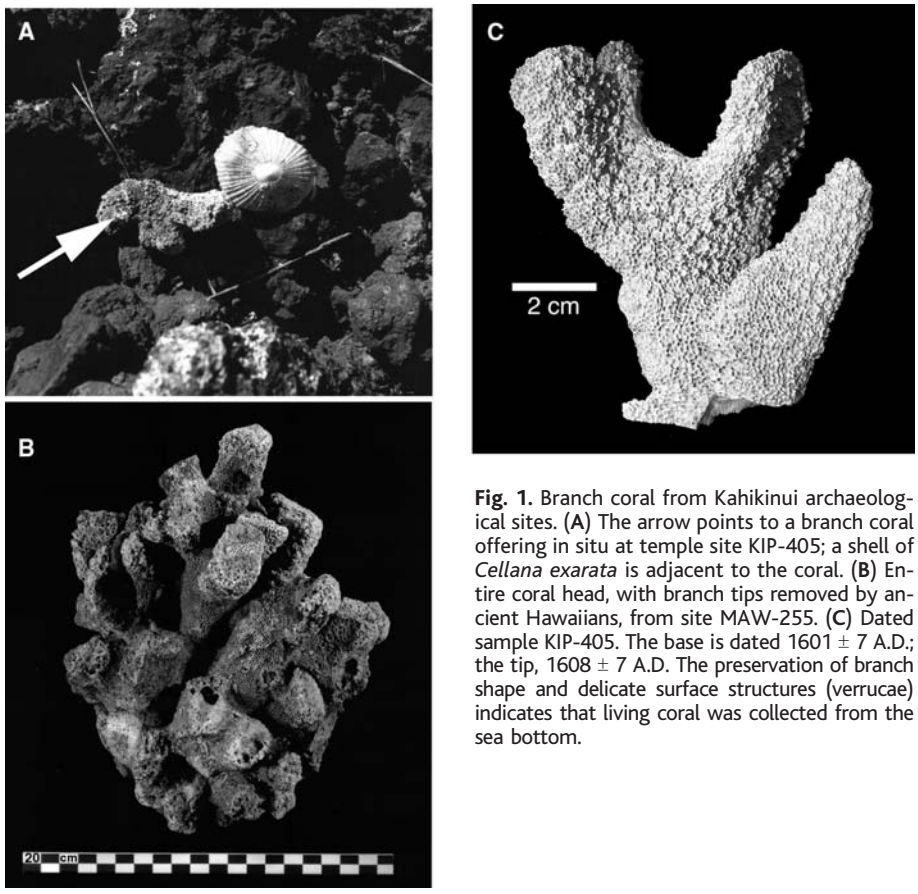


Fig. 1. Branch coral from Kahikinui archaeological sites. (**A**) The arrow points to a branch coral offering in situ at temple site KIP-405; a shell of *Cellana exarata* is adjacent to the coral. (**B**) Entire coral head, with branch tips removed by ancient Hawaiians, from site MAW-255. (**C**) Dated sample KIP-405. The base is dated 1601 ± 7 A.D.; the tip, 1608 ± 7 A.D. The preservation of branch shape and delicate surface structures (verrucae) indicates that living coral was collected from the sea bottom.

9 years. To encompass possible variability in $^{230}Th_{nr}$, we assigned an uncertainty of ±50% to the $^{230}Th_{nr}/^{232}Th$ ratio applied to fossil corals. The resulting $^{230}Th_{nr}$ correction adds <1 to 5 years to the calculated age errors, depending on the ^{232}Th content of each sample.

Calculated dates for the archaeological corals fall in a narrow range from 1565 ± 8 to 1638 ± 6 A.D. (Table 2). Moreover, dates on colony-base fragments may overestimate the harvest dates by as much as 30 to 40 years: the estimated longevity of individual *P. meandrina* colonies. If only the dates of coral branches and branch fragments from Kahikinui are considered, the age range is restricted to 1608 ± 7 to 1638 ± 6 A.D. Thus, we infer that all of the dated Kahikinui corals were placed as dedicatory offerings in an interval of ~60 years and perhaps in as little as 30 years—possibly within the span of a single generation of Hawaiians. Branches from the Kawela site on Molokai yield two analytically indistinguishable dates, 1571 ± 8 A.D. (weighted mean of two analyses of Kawela 1) and 1565 ± 8 A.D., which are slightly older than the Kahikinui branches.

Our sample of Kahikinui corals includes those from both small and mid-sized temples, as well as the largest temple structure (KIP-1010) in the district. Accordingly, the timing of intensive temple construction, reflecting a fundamental change in the sociopolitical structure of the district—the imposition of a ritual control hierarchy—occurred rapidly. The ages from Molokai indicate penecontemporaneous or perhaps slightly earlier construction of similar temples outside of Kahikinui.

Chiefly genealogies and oral traditions collected in the 19th century (*9*) provide detailed information about Hawaiian political history. For Maui Island, these traditions indicate that two formerly independent chiefdoms were brought under the control of a single leader during the reign of Pi'ilani, dated to approximately 1570–1600 A.D. Pi'ilani's grandson, Kamalalawalu (~1610–1630 A.D.), extended the Maui polity by taking over the nearby island of Lanai (and probably also smaller Kahoolawe Island). These conquests, coinciding with our dates for intensive temple construction, expanded the Maui polity in size from about 940 km^2 to more than 2360 km^2, the kind of territorial expansion predicted with the formation of an archaic state (*29*). A parallel process of sociopolitical change apparently occurred on Hawaii Island at the same time, under the reign of Liloa and his son 'Umi-a-Liloa (*9*).

The temples provide tangible archaeological evidence of the speed with which a fundamental sociopolitical transition occurred in proto-historic Hawaii. The surprising swiftness of the transition revealed by the new dates raises the possibility that similar transitions elsewhere may have been equally abrupt.

Table 1. Temple sites with dated branch coral samples.

Location	Site no.	Basal area (m^2)	Architectural form
Kahikinui, Maui	AUW-11	30	Walled enclosure
Kahikinui, Maui	KIP-273	174	Notched enclosure and platform
Kahikinui, Maui	KIP-275	120	Platform
Kahikinui, Maui	KIP-405	400	Notched enclosure and terrace
Kahikinui, Maui	KIP-414	433	Notched enclosure
Kahikinui, Maui	KIP-1010	1400	Double notched walled enclosures
Kahikinui, Maui	MAW-255	150	Terrace complex
Kawela, Molokai		90	Walled enclosure and platform

Table 2. ^{230}Th dates for archaeological and modern corals (95% confidence interval errors). Uncorrected dates assume no $^{230}Th_{nr}$, whereas corrected dates assume a $(^{230}Th_{nr}/^{232}Th)$ activity ratio of 2.5 ± 1.2 determined from analysis of a modern Kahikinui coral. Kawela 1-A and Kawela 1-B are replicate samples obtained by splitting 2 g of coral. ^{232}Th abundances are corrected for a total procedural blank of 29 ± 15 pg; the blank $(^{230}Th/^{232}Th)$ activity ratio is ≤3.7. Modern coral was collected in July 2002 and analyzed in December 2003. Isotopic ratios are activity ratios; decay constants used are those of (*30*). Full analytical data are available in the supporting online material (table S1).

Sample	Coral habit	^{232}Th (pg/g)	$(^{230}Th/^{232}Th)_{act}$	Uncorrected date (A.D.)	Corrected date (A.D.)
Kawela 1-A	Branch fragment	570	67.6	1557 ± 7	1575 ± 12
Kawela 1-B	Branch fragment	543	71.2	1552 ± 5	1569 ± 10
Kawela 2	Branch tip	297	106.4	1553 ± 6	1565 ± 8
AUW-11	Branch fragment	284	119.9	1629 ± 4	1638 ± 6
KIP-273	Branch tip	231	139.2	1610 ± 6	1618 ± 7
KIP-275	Branch fragment	194	133.5	1617 ± 5	1625 ± 6
MAW 255	Branch tip	350	108.1	1619 ± 5	1629 ± 7
KIP-405	Branch tip	360	91.7	1596 ± 4	1608 ± 7
KIP-405	8 cm from branch tip	249	147.2	1594 ± 6	1601 ± 7
KIP-414	Colony-base fragment	76	306.7	1569 ± 5	1574 ± 6
KIP-1010	Colony-base fragment	115	241.6	1574 ± 10	1580 ± 10
Modern	Branch tip	336	3.1	1992 ± 1	2002 ± 5

References and Notes

1. K. V. Flannery, *Annu. Rev. Ecol. Syst.* **3**, 399 (1972).
2. H. T. Wright, in *American Archaeology Past and Future*, D. J. Meltzer, D. D. Fowler, J. A. Sabloff, Eds. (Smithsonian Institution, Washington, DC, 1986), pp. 323–365.
3. K. V. Flannery, in *Archaic States*, G. M. Feinman, J. Marcus, Eds. (School of American Research, Santa Fe, NM, 1998), pp. 15–57.
4. M. D. Sahlins, *Stone Age Economics* (Aldine, Chicago, 1972).
5. P. V. Kirch, *The Evolution of the Polynesian Chiefdoms* (Cambridge Univ. Press, Cambridge, 1984).
6. P. V. Kirch, *Feathered Gods and Fishhooks* (Univ. of Hawaii Press, Honolulu, HI, 1985).
7. P. V. Kirch, *J. World Prehistory* **4**, 311 (1990).
8. R. Cordy, *Exalted Sits the Chief* (Mutual Publishing, Honolulu, HI, 2000).
9. C. K. C. Abad, thesis, Univ. of Hawaii, Honolulu, HI (2000).
10. V. Valeri, *Kingship and Sacrifice: Ritual and Society in Ancient Hawaii* (Univ. of Chicago Press, Chicago, 1985).
11. M. J. Kolb, *Asian Perspect.* **31**, 9 (1992).
12. M. J. Kolb, *Curr. Anthropol.* **35**, 521 (1994).
13. T. S. Dye, E. Komori, *N. Z. J. Archaeol.* **14**, 113 (1992).
14. J. Allen, *N. Z. J. Archaeol.* **14**, 45 (1992).
15. R. E. Taylor, *Radiocarbon Dating* (Academic Press, New York, 1987).
16. M. Stuiver *et al.*, *Radiocarbon* **40**, 1117 (1998).
17. R. L. Edwards, C. D. Gallup, H. Cheng, in *Uranium-Series Geochemistry*, B. Bourdon, G. M. Henderson, C. C. Lundstrom, S. P. Turner, Eds. (Mineralogical Society of America, Washington, DC, 2003), pp. 363–405.
18. J. S. Dean, in *Advances in Archaeological Method and Theory*, I. M. Schiffer, Ed. (Academic Press, New York, 1978), pp. 223–265.
19. This criterion does not hold for other kinds of corals, such as *Porites* spp., which were collected by Hawaiians from cobble beaches and used as abrading tools. In the latter case, the death age of the corals may significantly predate their use as tools.
20. P. V. Kirch, *Antiquity* **78**, 102 (2004).
21. P. V. Kirch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9936 (2004).
22. M. I. Weisler, P. V. Kirch, *N. Z. J. Archaeol.* **7**, 129 (1985).
23. Materials and methods and complete U-Th analytical data are available as supporting material on *Science* Online.
24. D. Gulko, *Hawaiian Coral Reef Ecology* (Mutual Publishing, Honolulu, HI, 1998).
25. R. L. Edwards, F. W. Taylor, G. J. Wasserburg, *Earth Planet. Sci. Lett.* **90**, 371 (1988).
26. K. M. Cobb, C. D. Charles, H. Cheng, M. Kastner, R. L. Edwards, *Earth Planet. Sci. Lett.* **210**, 91 (2003).
27. C. A. Huh, W. S. Moore, D. C. Kadko, *Geochim. Cosmochim. Acta* **53**, 1357 (1989).
28. M. Roy-Barman, J. H. Chen, G. J. Wasserburg, *Earth Planet. Sci. Lett.* **139**, 351 (1996).
29. C. S. Spencer, *Cult. Dyn.* **10**, 5 (1998).
30. H. Cheng, R. L. Edwards, M. T. Murrell, S. Goldstein, *Chem. Geol.* **169**, 17 (2000).

31. Supported by NSF grants BCS-0119819 and SBR-9805754 with P.V.K. as principal investigator and by the Ann and Gordon Getty Foundation. Permission to conduct archaeological research in Kahikinui was granted by the Department of Hawaiian Home Lands, State of Hawaii. For additional support and assistance, we thank Ka 'Ohana o Kahikinui, 'Ulupalakua Ranch, and the State of Hawai'i Historic Preservation Division. W.D.S. thanks K. Ludwig for discussion in the field and lab and G. Johnson for carrying out U-Th analyses.

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5706/102/DC1
Materials and Methods
Figs. S1 and S2
Table S1
References

20 September 2004; accepted 29 November 2004
10.1126/science.1105432

Genome Sequence of the PCE-Dechlorinating Bacterium *Dehalococcoides ethenogenes*

Rekha Seshadri,[1]* Lorenz Adrian,[2] Derrick E. Fouts,[1] Jonathan A. Eisen,[1,3] Adam M. Phillippy,[1] Barbara A. Methe,[1] Naomi L. Ward,[1,4] William C. Nelson,[1] Robert T. Deboy,[1] Hoda M. Khouri,[1] James F. Kolonay,[1] Robert J. Dodson,[1] Sean C. Daugherty,[1] Lauren M. Brinkac,[1] Steven A. Sullivan,[1] Ramana Madupu,[1] Karen E. Nelson,[1] Katherine H. Kang,[1] Marjorie Impraim,[1] Kevin Tran,[1] Jeffrey M. Robinson,[1] Heather A. Forberger,[1] Claire M. Fraser,[1,5] Stephen H. Zinder,[6] John F. Heidelberg[1,4]

Dehalococcoides ethenogenes is the only bacterium known to reductively dechlorinate the groundwater pollutants, tetrachloroethene (PCE) and trichloroethene, to ethene. Its 1,469,720–base pair chromosome contains large dynamic duplicated regions and integrated elements. Genes encoding 17 putative reductive dehalogenases, nearly all of which were adjacent to genes for transcription regulators, and five hydrogenase complexes were identified. These findings, plus a limited repertoire of other metabolic modes, indicate that *D. ethenogenes* is highly evolved to utilize halogenated organic compounds and H_2. Diversification of reductive dehalogenase functions appears to have been mediated by recent genetic exchange and amplification. Genome analysis provides insights into the organism's complex nutrient requirements and suggests that an ancestor was a nitrogen-fixing autotroph.

Tetrachloroethene (PCE) and trichloroethene (TCE) are among the most pervasive organic groundwater pollutants, primarily because of their disposal after use as industrial cleaners. Aerobic biodegradation of these solvents is ineffective; however, under anaerobic conditions, they can serve as terminal electron acceptors for dehalorespiration by bacteria that sequentially dechlorinate PCE to less chlorinated ethenes. Most cultured anaerobic (de)halorespirers (for example, *Sulfurospirillum multivorans* and *Dehalobacter restrictus*) only perform incomplete reductive dechlorination, yielding the toxic end product *cis*-dichloroethene (DCE) (*1*). In contrast, *Dehalococcoides ethenogenes* performs complete reductive dechlorination of PCE and TCE to the nontoxic metabolite ethene (*2*). *D. ethenogenes* belongs to subphylum 2 of *Chloroflexi* (green nonsulfur bacteria) (*3*–*5*), whereas other dehalorespiring bacteria are members of *Proteobacteria* and *Firmicutes* (*6*).

D. ethenogenes strain 195 was derived from an anaerobic sewage digestor. Subsequent studies demonstrated that *Dehalococcoides* spp. occur at many contaminated groundwater sites, and their presence is correlated with dechlorination past the product DCE (*5*, *7*–*9*). *Dehalococcoides* spp. have also been shown to dechlorinate chlorobenzenes, chloronaphthalenes, polychlorinated biphenyls, and dibenzodioxins (*4*, *10*, *11*). *D. ethenogenes* exhibits an unusual metabolic specialization, using only H_2 as an electron donor and chlorinated compounds as electron acceptors to support growth (*2*). In culture, it also requires acetate, vitamin B_{12}, and extracts of mixed microbial cultures as nutrients, suggesting a relatively limited repertoire of biosynthetic capabilities.

The *D. ethenogenes* genome sequence provides a better understanding of dehalorespiration and its evolution. Because the organism is difficult to culture and largely intractable to study using contemporary molecular genetic methods, the genome sequence contributes to our understanding of the physiology of *D. ethenogenes* and the evolution of catabolic pathways.

The genome of *D. ethenogenes* is composed of a 1,469,720–base pair (bp) circular chromosome containing 1591 predicted coding sequences (CDSs) (table S1). The genome possesses large duplicated regions and several integrated elements (Table 1 and fig. S1). A 31-kb tandem duplication encodes enzymes for CO_2 fixation, corrinoid cofactor salvage, and transport, among others (fig. S2). Within this duplication, DET0674 may be a fusion of DET0708 (aspartyl–transfer RNA synthetase) and DET0640 (universal stress protein family). With the exception of a single frameshift in a histidine kinase gene (DET0696), no nucleotide polymorphisms were detected between the tandem duplications. The mechanism that generates this duplication is unclear because flanking regions have no repeats; however, nonhomologous illegitimate recombination (*12*) may have occurred. Although this duplication could represent a strain-specific event that may not prevail in the population, it seems likely that a selective advantage is conferred (*13*–*15*); for example, by increasing gene dosage, particularly for cofactor salvage.

A ~22-kb integrated element (IE), which contains phage-like genes and is flanked by putative attachment (att) sites, occurs in at least three copies (fig. S3 and Table 1) as a tandem duplication (elements III and IV) and a distal third copy (element VI). A very large number of clones correspond to the current assembly. However, because of the high redundancy of sequence data, the presence of additional tandem copies or episomal elements cannot be ruled out. These and other IEs (Table 1) represent ~13.6% of the genome of *D. ethenogenes*.

It has been speculated that anthropogenic compounds select for microbial populations that have acquired the ability to use them (*16*). Only in the past 50 years have substantial quantities of chlorinated ethene solvents been released into the environment. It appears that the amplification of dechlorination capabilities in this organism has been mediated by recent genetic exchange. For instance, four of the putative IEs (Table 1), which have atypical trinucleotide composition and encode site-specific recombinases, contain reductive dehalogenase (RD) genes. This suggests their recent acquisition (or perhaps consignment for dissemination). The gene that encodes TceA-RD, which essentially defines this organism's ability to dechlorinate chloroethenes past DCE (*17*), is located in an IE.

Previous studies have shown that *D. ethenogenes* reduces PCE to ethene using two RDs (*18*) belonging to a family of iron-sulfur corrinoid proteins (*19*, *20*). One of the RDs reduces PCE to TCE, and a second RD(TceA) reduces TCE and other chloroethenes to vinyl chloride and ethene (*17*). The genome sequence indicates 17 intact CDSs (table S2) that encode putative RDs (*21*), all possessing potential twin-arginine transport (TAT) signals for export to the periplasm

[1]The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA. [2]Fachgebiet Technische Biochemie, Institut für Biotechnologie, Technische Universität Berlin, Seestrasse 13, D-13353 Berlin, Germany. [3]Johns Hopkins University, Charles and 34th Streets, Baltimore, MD 21218, USA. [4]Center of Marine Biotechnology, University of Maryland Biotechnology Institute, Baltimore, MD 21202, USA. [5]Departments of Pharmacology and Microbiology and Tropical Medicine, George Washington University School of Medicine, 2300 Eye Street N.W., Washington, DC 20037, USA. [6]Section of Microbiology, Cornell University, 272 Wing Hall, Ithaca, NY 14853-5701, USA.

*To whom correspondence should be addressed. E-mail: rekha@tigr.org

(*22*), with adjacent genes that encode potential membrane-anchoring proteins. Additionally, two degenerate RD genes are present. One (DET0088) is truncated and lacks an adjacent anchoring protein, and the other (DET0162) has a point mutation. The membrane anchor (DET0163) of DET0162 is shorter, and flanking transposases (DET0165 and DET0166) suggest ongoing decay. Seventeen RD genes are located within 300,000 bp of the predicted origin of replication, and many are clustered (fig. S1); only DET1171 and DET0876 are located distally. With the exception of DET0079, which encodes TceA-RD, all RD operons are oriented in the direction of replication. Such a strong orientation bias was not seen with other gene groups.

D. ethenogenes has been shown to dehalogenate numerous halogenated substrates (*4*, *10*, *11*). Given the presence of multiple potential RDs, it is likely that some may have specificities for these alternate compounds. The *D. ethenogenes* RD genes are part of a gene family that includes confirmed and putative RDs from other dehalogenating organisms, sequences from uncultured Sargasso Sea microorganisms (*23*), and the *Silicibacter pomeroyi* genome (fig. S4).

With one exception (DET0079, *tceA*), all intact RD genes are found in close proximity to genes for transcription regulators, particularly two-component signal transduction systems (TCSs) (table S2), which suggests stringent regulation of RD activity. The RD genes with no proximal TCS (DET0079, DET0876, and DET0162) are located within putative integrated regions (Table 1). Typically, sensor histidine kinase (HK) components of TCS possess at least two transmembrane regions and function as periplasmic membrane receptors that detect environmental signals. However, all of the RD-associated HKs lack transmembrane helices, suggesting that these are soluble cytoplasmic proteins (*24*). In contrast, many other HK components are more typical, possessing two to four transmembrane regions (table S3). Phylogenetic analysis of these HK components suggests a major expansion of at least one group of HKs that are associated with RD genes (fig. S5).

The predicted cytoplasmic location suggests that these sensor HKs respond to intracellular rather than extracellular stimuli. One speculation is that they may detect redox levels of cytoplasmic electron carriers, based on the presence of ligand-binding PAS/PAC motifs (table S3). These electron carriers have been implicated in sensing cellular energy reserves through changes in proton motive force (PMF) or redox potential of the electron transport chain (*25*). This sensing of energy reserves would allow the cell to respond rapidly to changes in energy status, which is a critical function for an organism that has limited means for energy production. Another speculation is that soluble HKs could detect nonpolar chlorinated substrates that permeate the cell membranes, resulting in up-regulation of RD expression when the substrate becomes available.

The genome encodes five different putative membrane-bound multisubunit hydrogenases (Fig. 1 and table S4) (*26*), which are typically involved in energy transduction. Of these, only Hup hydrogenase possesses a TAT signal in the small catalytic subunit and is thus predicted to occur in the periplasm. This location makes Hup hydrogenase a good candidate as an uptake hydrogenase involved in energy conservation, because protons produced from H_2 oxidation will be outside the cell and will not dissipate a PMF (Fig. 1). The Ech hydrogenase is known to be a proton pump, and it can use a PMF to drive electrons thermodynamically uphill (reverse electron transport) (*27*). This uphill drive generates low-potential electrons that are needed to either reduce cobalt in the RDs to the Co^{+1} state (*28*) or for biosynthetic reactions. Including those involved in hydrogenase maturation, over 40 genes in *D. ethenogenes* are potentially devoted to the simple reaction that interconverts H_2 with protons and electrons, signifying the importance of this reaction.

An 11-CDS operon encoding reduced nicotinamide adenine dinucleotide (NADH)–ubiquinone oxidoreductase (complex I, Nuo) lacks genes for a cytoplasmic input module (NuoEFG) that withdraws electrons from NAD(P)H. Instead, these genes are found associated with other oxidoreductase operons (table S4). It is not clear whether one of the distally located NuoEFG modules or a putative cofactor F_{420} input module (FpoF) serves as the input domain of the *D. ethenogenes* Nuo-assembly, as described for Complex I assemblies of *Methanosarcina mazei* (*29*). Thus, it is difficult to predict the electron donor for this complex, and it is not clear whether the complex uses quinone as its electron acceptor.

The only operon potentially involved in the reduction of nonhalogenated electron acceptors is a putative membrane-bound molybdopterin-containing oxidoreductase. The molybdopterin-containing α subunit has a predicted periplasmic location and shares 28% identity with tetrathionate reductase from *Salmonella typhimurium*, although this low level of identity makes it difficult to predict its substrate.

D. ethenogenes has not been found to use formate as an electron donor (*2*); however, a putative membrane-bound formate dehydrogenase was detected. The methanogenic Archaeae *Methanocaldococcus jannaschii* and *M. thermoautrophicus* use only H_2 as an electron donor, but both possess genes homologous to those encoding formate dehydrogenases (*30*). If not involved in electron transport, the formate dehydrogenase may be involved in producing formate for biosynthesis.

Table 1. *D. ethenogenes* strain 195 putative IEs. Putative att sites were determined using MUMmer or Grasta software. The best matches were determined by Basic Local Alignment Search Tool (BLASTP) searches of a limited custom phage database. The putative attL/R sequence included here includes only the att site closest to recombinase.

Integrated element	Region	Locus names	Size (bp)	Target for integration	putative attL/R	Atypical trinucleotide?	%GC	Total # ORFS	# phage BLASTP hits	# conserved hypotheticals	# hypotheticals	Phage integrase	Resolvase	restriction endonuclease	methylase	DNA polymerase	DNA primase	Helicase	Transposase	*gin* -like invertase	*lexA* -like repressor	Reductive Dehalogenase
I	57850-88645	DET0063-0091	30,796	tRNA-Ala-1	ctccggctc	yes	46.7	28	15	2	11		2		1	1	1	1	1	1		DET0079
II	149104-165172	DET0155-0169	16,069	tRNA-Val-1	cgcccacca	yes	47.6	14	5	2	6	1	1			1			2	1		DET0162
III	240759-263123	DET0251-0272	22,366	TCTGATT	TCTGATT	yes	45.7	22	6	1	15	1	1				1			1	1	
IV	263124-285490	DET0273-0295	22,366	pstRNA-Lys	TAGGCCGTACAGGTCTCGAACCTGTGACCCTCTGATT	yes	45.7	23	6	1	16	1	1				1			1	1	
V	803127-810526	DET0875-0883	7,400	tRNA-Lys-2	TAGGCCGTACAGGTCTCGAACCTGTGACCCTCTGATTAAAAG	yes	38.9	9	2	3	2	1										DET0876
VI	810529-832902	DET0884-0905	22,366	tRNA-Lys-2	TAGGCCGTACAGGTCTCGAACCTGTGACCCTCTGATTAAAAGTCA	yes	45.7	22	6	1	15	1	1				1			1	1	
VII	968488-1023745	DET1066-1118	55,258	unknown	unresolved	yes	50.6	52*	31	11	23	3		2	2	1		1				
VIII	1325415-1338416	DET1472-1478	13,002	tRNA-Val-3	TTTACACGCAGGGGGTCATAGGTTCGAATCCTATACCGCCCACCA	yes	39.3	7	1	0	4	1										
IX	1398183-1407943	DET1552-1561	9,761	tRNA-Ala-3	CTCAGCTCCA	yes	40.8	10	2	1	3		1									DET1559
Total amount of DNA:			199,384																			
% of chromosome:			13.6%																			

*Region VII shows extensive synteny with the *Bacillus cereus* 10987 prophage.

Observations of enhanced growth in mixed culture (stimulated by amino acids and cell extracts), as opposed to poor growth of pure cultures of *D. ethenogenes*, suggest a dependent ecological interaction (symbiosis) with other species in an anaerobic consortium, as suggested previously (*31*). Aspects of the *D. ethenogenes* metabolic profile support this notion.

D. ethenogenes requires acetate as a carbon source (*2*). Acetyl–coenzyme A (CoA) synthetase, pyruvate-ferredoxin oxidoreductase, and gluconeogenesis enzymes indicate a pathway for acetate assimilation (fig. S6A). The tricarboxylic acid (TCA) cycle in anaerobes is usually incomplete and provides precursors for amino acid synthesis. Some genes encoding TCA-cycle enzymes were found in a single operon; however, the enzyme for a key step, citrate synthetase, which is needed to form glutamate family amino acids, is absent.

In the previously mentioned 31-kb tandem duplication, several CDSs involved in the Wood/Ljungdahl acetyl-CoA pathway of CO_2 fixation (*32*) were predicted. These include formyltetrahydrofolate ligase, a corrinoid iron-sulfur protein, and the α subunit of carbon monoxide dehydrogenase/acetyl-CoA decarbonylase-synthase (CODH/ACDS). Other essential components, such as methylene-tetrahydrofolate reductase and the β subunit of CODH/ACDS responsible for CO_2 reduction to CO (*32*), are missing (fig. S6B). Perhaps an ancestor of *D. ethenogenes* was an autotroph using the acetyl-CoA pathway, which may have been disrupted after genetic rearrangements. Phylogenomic analysis strongly supports the classification of *D. ethenogenes* as a relative of the green nonsulfur bacteria, all of which are photoautotrophic.

D. ethenogenes possesses a nitrogenase-encoding operon (*nifHI$_1$I$_2$DKENB*, DET1151-1158) and a distal gene, *nifV* (DET1614), which encodes homocitrate synthetase used in nitrogenase FeMo-cofactor biosynthesis. Phylogenetically, the nitrogenase structural genes belong to cluster III (*33*), which includes nitrogenases from diverse anaerobic Bacteria and Archaea. The *nif* cluster is adjacent to an adenosine triphosphate (ATP)–binding cassette (ABC) transporter for molybdenum, an element essential to nitrogenase. These findings suggest that *D. ethenogenes* can fix nitrogen, which is difficult to demonstrate in the undefined complex medium in which this organism grows and is surprising, in light of the bacterium's complex nutrient requirements.

Complete pathways for the biosynthesis of all of the amino acids were identified, with the exception of methionine and possibly glutamate. Pathways for the synthesis of some cofactors appear incomplete. *D. ethenogenes* requires the corrinoid vitamin B_{12} in large amounts for growth (*2*). Corrinoids serve as a prosthetic group in bacterial RDs, homologous to those from *D. ethenogenes* (*34*), and inhibitor studies implicate corrinoids in *D. ethenogenes* RDs (*18*). Although no de novo corrinoid-ring biosynthetic genes are encoded, the 31-kb duplication (fig. S2) contains several genes (*cobCDSTU*) for corrinoid salvage (*35*) and an adjacent cobalamin/$Fe3^+$ ABC transporter. Amplification of these genes may have allowed for more efficient uptake and salvage of exogenously obtained corrin precursors. Because there is little nucleotide difference between the duplicates, the repeat appears quite recent and may have occurred during the period when the organism was cultured in the absence of added vitamin B_{12} (*2*).

Similarly, genes for quinone modification, but not for quinone-ring synthesis, were found, although the role of quinones in electron transport in *D. ethenogenes* is unclear. Instead of genes for biotin synthesis, a puta-

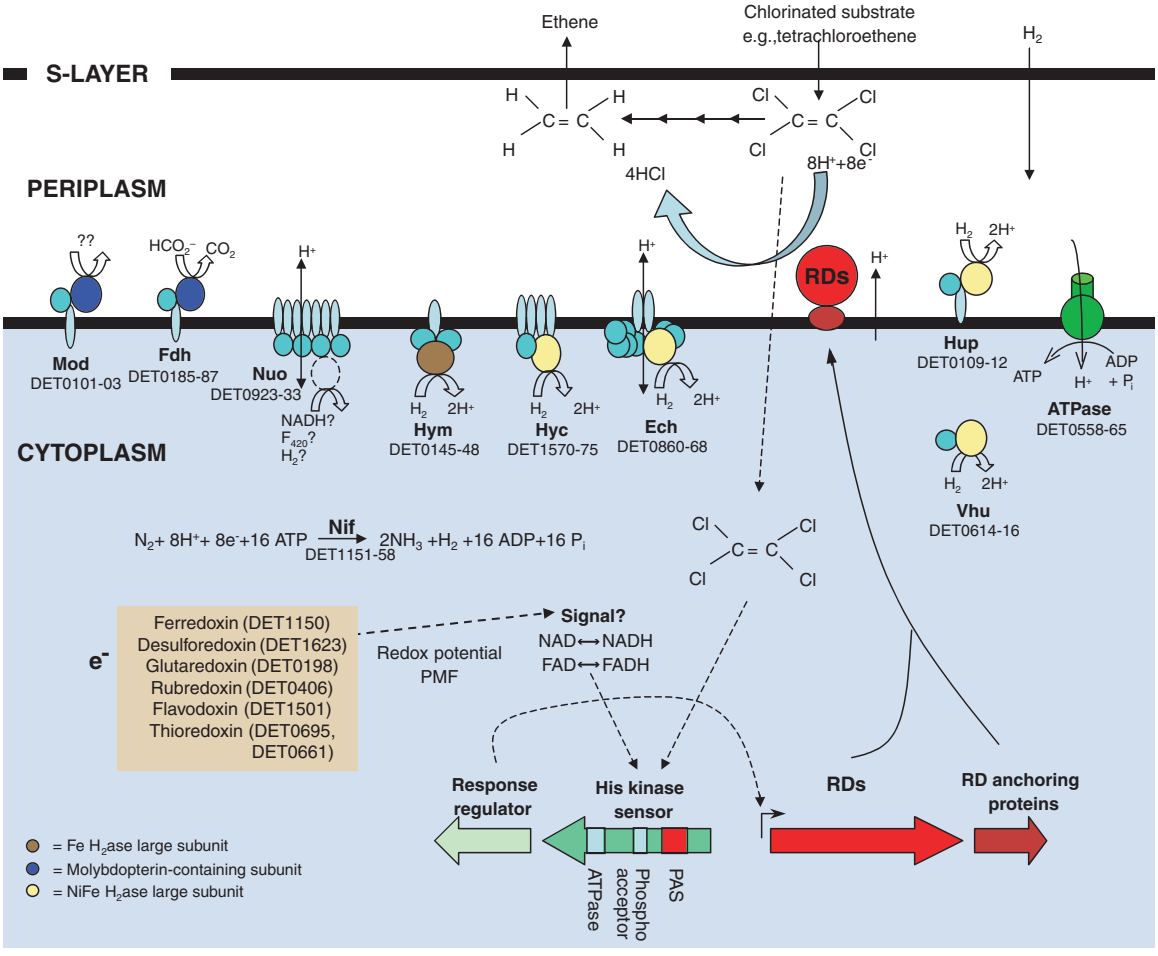


Fig. 1. Schematic representing the enzymes and regulatory network hypothesized from genome analysis to control dechlorination activities by *D. ethenogenes*. This network involves cytoplasmic two-component system regulators that sense various stimuli and up-regulate reductive dehalogenase activity in response to the redox status of the cell. H_2ase, hydrogenase. For a listing of loci encoding putative reductive dehalogenases and two-component systems, see tables S2 and S3, respectively.

tive *bioY* (DET1184) encoding a biotin transporter was found.

Thus, the genome sequence provides insight into the organism's complex nutrient requirements and its commitment to the dehalorespiratory process. The genome also suggests that an ancestor was a nitrogen-fixing autotroph. In the long term, genome data will serve as a foundation for the development of phylogenetic and functional marker probes, for detection and monitoring of *D. ethenogenes* in the environment, and for population genetic studies.

References and Notes

1. H. Smidt, W. M. De Vos, *Annu. Rev. Microbiol.* **58**, 43 (2004).
2. X. Maymó-Gatell, Y. Chien, J. M. Gossett, S. H. Zinder, *Science* **276**, 1568 (1997).
3. P. Hugenholtz, B. M. Goebel, N. R. Pace, *J. Bacteriol.* **180**, 4765 (1998).
4. L. Adrian, U. Szewzyk, J. Wecke, H. Görisch, *Nature* **408**, 580 (2000).
5. D. E. Fennell, A. B. Carroll, J. M. Gossett, S. H. Zinder, *Environ. Sci. Technol.* **35**, 1830 (2001).
6. S. H. Zinder, in *The Encyclopedia of Environmental Microbiology*, G. Bitton, Ed. (Wiley, New York, 2002), pp. 507–516.
7. F. E. Löffler, Q. Sun, J. Li, J. M. Tiedje, *Appl. Environ. Microbiol.* **66**, 1369 (2000).
8. E. R. Hendrickson *et al.*, *Appl. Environ. Microbiol.* **68**, 485 (2002).
9. D. W. Major *et al.*, *Environ. Sci. Technol.* **36**, 5106 (2002).
10. J. He, K. M. Ritalahti, K. L. Yang, S. S. Koenigsberg, F. E. Loffler, *Nature* **424**, 62 (2003).
11. D. E. Fennell, I. Nijenhuis, S. F. Wilson, S. H. Zinder, M. M. Häggblom, *Environ. Sci. Technol.* **38**, 2075 (2004).
12. S. D. Ehrlich *et al.*, *Gene* **135**, 161 (1993).
13. A. B. Reams, E. L. Neidle, *Annu. Rev. Microbiol.* **58**, 119 (2004).
14. A. B. Reams, E. L. Neidle, *Mol. Microbiol.* **47**, 1291 (2003).
15. A. B. Reams, E. L. Neidle, *J. Mol. Biol.* **338**, 643 (2004).
16. J. R. van der Meer, V. Sentchilo, *Curr. Opin. Biotechnol.* **14**, 248 (2003).
17. J. K. Magnuson, M. F. Romine, D. R. Burris, M. T. Kingsley, *Appl. Environ. Microbiol.* **66**, 5141 (2000).
18. J. K. Magnuson, R. V. Stern, J. M. Gossett, S. H. Zinder, D. R. Burris, *Appl. Environ. Microbiol.* **64**, 1270 (1998).
19. A. Neumann, H. Scholz-Muramatsu, G. Dieckert, *Arch. Microbiol.* **162**, 295 (1994).
20. B. A. van de Pas *et al.*, *J. Biol. Chem.* **274**, 20287 (1999).
21. R. Villemur, M. Saucier, A. Gauthier, R. Beaudet, *Can. J. Microbiol.* **48**, 697 (2002).
22. F. Sargent, *Trends Microbiol.* **9**, 196 (2001).
23. J. C. Venter *et al.*, *Science* **304**, 66 (2004).
24. M. Schobert, H. Gorisch, *Microbiol.* **147**, 363 (2001).
25. C. Wagner *et al.*, *Infect. Immun.* **70**, 6121 (2002).
26. P. M. Vignais, B. Billoud, J. Meyer, *FEMS Microbiol. Rev.* **25**, 455 (2001).
27. J. Meuer, H. C. Kuettner, J. K. Zhang, R. Hedderich, W. W. Metcalf, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5632 (2002).
28. W. Schumacher, C. Holliger, A. J. Zehnder, W. R. Hagen, *FEBS Lett.* **409**, 421 (1997).
29. S. Bäumer *et al.*, *J. Biol. Chem.* **275**, 17968 (2000).
30. D. R. Smith *et al.*, *J. Bacteriol.* **179**, 7135 (1997).
31. T. D. DiStefano, J. M. Gossett, S. H. Zinder, *Appl. Environ. Microbiol.* **58**, 3622 (1992).
32. T. I. Doukov, T. M. Iverson, J. Seravalli, S. W. Ragsdale, C. L. Drennan, *Science* **298**, 567 (2002).
33. J. P. Zehr, B. D. Jenkins, S. M. Short, G. F. Steward, *Environ. Microbiol.* **5**, 539 (2003).
34. A. Neumann *et al.*, *Arch. Microbiol.* **177**, 420 (2002).
35. M. J. Warren, E. Raux, H. L. Schubert, J. C. Escalante-Semerena, *Nat. Prod. Rep.* **19**, 390 (2002).

36. The complete genome sequence reported in this paper has been submitted to GenBank (accession number d_ethenogenes_195_68 CP000027). Supported by the U.S. Department of Energy, Office of Biological Energy Research, cooperative agreement DE-FC02-95ER61962. L.A. was supported by NSF grant MCB 0236044. We thank G. S. A. Myers, S. Salzberg, M. Pop, O. White, M. Heaney, S. Lo, M. Holmes, M. Covarrubias, J. Sitz, A. Resnick, J. Zhao, M. Zhurkin, S. Sengamalay, P. Sellers, R. Deal, R. Karamchedu, and V. Sapiro for informatics, database, and software support. Sequence data for *Silicibacter pomeroyi* were obtained from The Institute for Genomic Research Web site (www.tigr.org).

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5706/105/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S4
References

1 July 2004; accepted 4 November 2004
10.1126/science.1102226

Y Chromosome of *D. pseudoobscura* Is Not Homologous to the Ancestral *Drosophila* Y

Antonio Bernardo Carvalho[1]* and Andrew G. Clark[2]

We report a genome-wide search of Y-linked genes in *Drosophila pseudoobscura*. All six identifiable orthologs of the *D. melanogaster* Y-linked genes have autosomal inheritance in *D. pseudoobscura*. Four orthologs were investigated in detail and proved to be Y-linked in *D. guanche* and *D. bifasciata*, which shows that less than 18 million years ago the ancestral *Drosophila* Y chromosome was translocated to an autosome in the *D. pseudoobscura* lineage. We found 15 genes and pseudogenes in the current Y of *D. pseudoobscura*, and none are shared with the *D. melanogaster* Y. Hence, the Y chromosome in the *D. pseudoobscura* lineage appears to have arisen de novo and is not homologous to the *D. melanogaster* Y.

The origin and evolution of the *Drosophila* Y seem to be different from that of the canonical (e.g., mammalian) Y chromosomes. Mammalian sex chromosomes originated from an ordinary pair of autosomes, so that when one of the homologs acquired a strong male-determining gene and became a Y, the other homolog became the X. Progressive gene loss from the Y resulted in a mostly degenerated chromosome (*1*, *2*). The main evidence for this common origin of the sex chromosomes is that most of the mammalian Y-linked genes and pseudogenes are shared with the X (*3*). On the other hand, in *D. melanogaster* none of the nine known single-copy Y-linked genes have homologs on the X. Instead, their closest homologs are autosomal (*4*, *5*), which strongly suggests that they were acquired from the autosomes by gene duplications, as has been shown for two mammalian Y genes (*3*). *D. melanogaster* X and Y chromosomes do share two multicopy genes, but it is doubtful that they represent common ancestry: *Ste-Su*(*Ste*) genes were recently acquired from an autosome (*6*), whereas *rDNA* genes are present in nonhomologous chromosomes in different *Drosophila* species (*5*, *7*, *8*). The lack of any clear sign of X-Y homology is consistent with the hypothesis that the *Drosophila* Y did not originate from the degeneration of an X-homolog, but rather from a supernumerary ("B") chromosome that acquired X-pairing and male-related genes, though it is also possible that the degeneration went so far that all signs of homology were erased (*5*, *9*). Whatever its true origin, the *melanogaster*-like *Drosophila* Y is at least 63 million years (My) old, dating back to the separation of the Sophophora and *Drosophila* subgenera (*10*), because at least three genes (*kl-2*, *kl-3*, and *kl-5*) are shared between the Y chromosomes of *D. melanogaster* (*4*, *5*) and *D. hydei* (*11*, *12*).

The assembled *D. pseudoobscura* genome sequence (*13*) now makes possible a genome-wide study of Y-linked genes in a second *Drosophila* species. Given that this species also belongs to the subgenus Sophophora, we expected it to share the ancestral Y chromosome with *D. melanogaster*, as does the more distant *D. hydei*. In fact, we found in the *D. pseudoobscura* genome orthologs for most *D. melanogaster* Y-linked genes (*kl-2*, *kl-3*, *kl-5*, *ORY*, *PPr-Y*, and *ARY*), as well as the orthologs of their autosomal parental genes (*14*). The orthology is strongly supported by phylogenetic analysis (Fig. 1 and fig. S1) and by the reciprocal best match criteria (*14*). However, when we tested for Y linkage by polymerase chain reaction (PCR) with genomic DNA from males and females, we found that all six genes are present in both sexes, ruling out Y linkage. A

[1]Departamento de Genética, Universidade Federal do Rio de Janeiro, Caixa Postal 68011, CEP 21944-970, Rio de Janeiro, Brazil. [2]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA.

*To whom correspondence should be addressed. E-mail: bernardo@biologia.ufrj.br

formal genetic analysis by means of single-nucleotide polymorphisms that exist between the subspecies *D. p. pseudoobscura* (the sequenced one) and *D. p. bogotana* demonstrated that all six genes have autosomal inheritance in *D. pseudoobscura* (fig. S2). Thus, the ancestral *Drosophila* Y chromosome suffered at least one translocation to an autosome in the *D. pseudoobscura* lineage (supporting online text). In this study, we analyzed in detail the orthologs of *kl-2*, *kl-3*, *ORY*, and *PPr-Y* genes; *ARY* and *kl-5* will be dealt with in a separate paper. We failed to find the orthologs of four genes (*14*).

To date the Y → autosome translocation event(s) we examined progressively more distantly related species (*15*), by doing PCR in males and females with the same primers used for *D. pseudoobscura*. As shown in Fig. 2, we found that the four genes (*kl-2*, *kl-3*, *ORY*, and *PPr-Y*) are present in both sexes of *D. persimilis* and *D. miranda* (pseudoobscura subgroup) and of *D. affinis* and *D. azteca* [affinis subgroup; 2- to 3-My ago divergence from *D. pseudoobscura* (*16*)], and are male-specific in *D. guanche* and *D. bifasciata* [obscura subgroup; 18-My ago divergence (*10*)]. This shows that the Y-autosome translocation(s) occurred between 2 and 18 My ago [these numbers are approximations, because the divergence times are not precisely known (*10*)]. These results also help explain the observation that, in contrast to almost all *Drosophila* species, *D. affinis* males devoid of Y chromosomes are fertile (*17*).

The *D. pseudoobscura* genes described here almost certainly are functional, because (i) the coding regions have no premature stop codons, (ii) the putative splice junctions have the correct sequence, and (iii) reverse transcriptase (RT)–PCR shows that all genes are transcribed and are correctly spliced (fig. S3). Most of these genes have testis-restricted expression, as do their Y-linked orthologs in *D. melanogaster* (fig. S3) (*4*, *5*).

The Y-autosome translocation(s) seems to have led to a profound genomic reorganization in the former Y-linked genes. In *D. melanogaster* and *D. hydei* the Y-linked genes are unusually large [~1 to 5 megabase pairs (Mbp)], owing to megabase-sized introns (*11*, *18*). Intergenic distances also seem to be in the Mbp range. We infer that in the ancestral state, these genes had huge introns and large intergenic spacing. In *D. pseudoobscura*, *kl-2* and *kl-3* span a mere 28 and 105 kb (the cDNAs themselves each span 13 kb), which strongly suggests that they shrank at least to 1/10th of their original size after the Y-autosome translocation (fig. S4). Furthermore, *kl-2* and *ORY* are separated by 20 kb, and *ORY* and *PPr-Y* by 15 kb, whereas in no case was any pair of Y-linked genes in *D. melanogaster* even on the same scaffold. These observations strongly suggest that intergenic distances were also reduced.

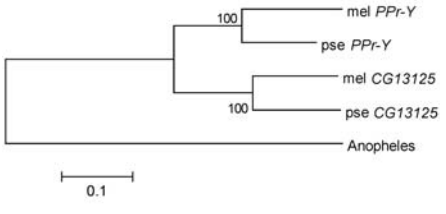


Fig. 1. Phylogeny of *PPr-Y* and its parental gene *CG13125*. Orthology between the *PPr-Y* genes of *D. melanogaster* and *D. pseudoobscura* is indicated by their statistically significant grouping (note the bootstrap test). The same result was obtained with *ORY*, *kl-2*, *kl-3*, *kl-5*, and *ARY* (fig. S1). The bar indicates the number of amino acid substitutions per site. *D. melanogaster PPr-Y* is Y-linked, and all other genes are autosomal. Accession numbers: mel *CG13125*, NP_609325; mel *PPr-Y*, AAL25119; and *Anopheles CG13125* ortholog, XP_319283. The sequences of *D. melanogaster ARY* and all *D. pseudoobscura* genes are available in the Third Party Annotation Section of the DDBJ/EMBL/GenBank databases under the accession numbers TPA: BK005622–BK005633 and BK005606.

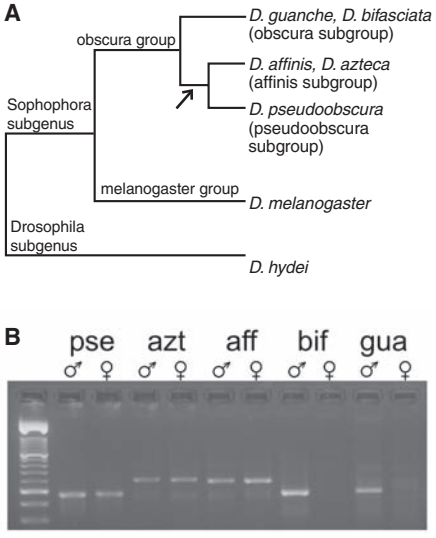


Fig. 2. Phylogenetic origin of the Y-autosome translocation. (**A**) Simplified phylogeny of the genus *Drosophila* (*15*). (**B**) PCR amplification of the *kl-3* gene in males and females of *D. pseudoobscura* (pse, lanes 2 and 3), *D. azteca* (azt, lanes 4 and 5), *D. affinis* (aff, lanes 6 and 7), *D. bifasciata* (bif, lanes 8 and 9), and *D. guanche* (gua, lanes 10 and 11). Y-linkage only occurs in the last two species, which belong to the more distant obscura subgroup. Amplification of *kl-3* in both sexes also occurred in *D. persimilis* and *D. miranda* (pseudoobscura subgroup; not shown). The same pattern of *kl-3* was observed for *kl-2*, *PPr-Y*, and *ORY*. These data show that the Y-autosome translocation occurred after the split of the obscura subgroup, and before the affinis-pseudoobscura split [arrow in (A)]. The identity of the PCR products was verified by sequencing (GenBank accession numbers: AY808044 to AY808058). Primer sequences and PCR conditions are available upon request.

At least in genomic structure, the Y genes appear to have become nearly euchromatic after the translocation to the autosome.

D. pseudoobscura does have a free Y chromosome, and given the randomness of the whole-genome shotgun sequencing strategy, portions of its sequence must be among the unmapped scaffolds (*19*). We searched for Y-linked genes using the fact that the whole-genome shotgun libraries were prepared from unsexed embryos, and hence Y-linked DNA should have about one-fourth as many reads as autosomal DNA (*19*). This and other computational methods yielded "candidates" that were then tested for Y linkage by PCR (fig. S5). Among 37 tested scaffolds, we identified 14 Y-linked ones (~150 kb) that contain 15 genes and pseudogenes (Table 1; tables S1 and S2). All scaffolds had a close euchromatic counterpart (>90% nucleotide identity); they arose through three segmental duplications involving several genes, one small genomic duplication, and one retrotransposition (table S1). None of these are shared with the *D. melanogaster* Y (Table 1 and table S1), which further supports the conclusion that these chromosomes are not homologous. However, it remains possible that the large uncharacterized parts of the two Y chromosomes are homologous. Efforts to identify more Y-chromosomal genes in both species and infer their functions are under way. The evidence that the 15 Y-linked genes and pseudogenes of *D. pseudoobscura* are not Y-linked in *D. melanogaster* is the following. BlastN searches using the *D. melanogaster* coding sequences of these genes and pseudogenes as the query and the *D. melanogaster* genome as the database al-

Table 1. Gene content of the *D. melanogaster* and *D. pseudoobscura* Y chromosomes.

Only in *D. melanogaster* Y*	Shared genes	Only in *D. pseudoobscura* Y†
kl-2	None found	*CG12218Y*-Ψ (X)
kl-3		*CG12848Y* (2R)
kl-5		*CG3894Y*-Ψ (2R)
ORY		*CG3880Y*-Ψ (2R)
PPr-Y		*CG10274Y* (3L)
ARY		*CG10289Y*-Ψ (3L)
Pp1-Y1		*CG7376Y*-Ψ (3L)
Pp1-Y2		*CG10191Y* (3L)
PRY		*CG17687Y*-Ψ (3L)
CCY		*CG32120Y*-Ψ (3L)
		CG10171Y (3L)
		CG14110Y (3L)
		CG14111Y (3L)
		CG14112Y (3L)
		CG31175Y-Ψ (3R)

*Orthologs in the *D. pseudoobscura* genome were detected for the first six genes, but these are located in an autosome. No ortholog was found for the last four genes. †In *D. pseudoobscura* all 15 genes are present as duplications from euchromatic genes. In *D. melanogaster* only the euchromatic copy was found, whose location is shown in parentheses. Pseudogenes were assigned with the standard Ψ symbol. See also table S1.

ways detected only the euchromatic copy, whereas the same search against the *D. pseudoobscura* genome readily detected two or more copies (one euchromatic and at least one Y-linked copy). Sequence gaps in the *D. melanogaster* Y chromosome cannot explain this result: Although the assembly of the Y chromosome is disrupted in the highly repetitive sequences (*19*), the coding sequences are fairly well represented (table S1).

An explanation for this exceptional Y-chromosome turnover may lie in the X chromosome of the *D. pseudoobscura* lineage. The *D. pseudoobscura* X originated from a fusion of the ancestral X and an autosome [the homolog of the *D. melanogaster* 3L chromosome, also called the Muller D element (Fig. 3) (*13*, *20*)]. This fusion and the Y-autosome translocation(s) co-occur along the phylogeny: Both are present in the subgroups pseudoobscura and affinis, and both are absent in the obscura subgroup. In these X-autosome fusion systems, one member of the autosomal pair became part of the X and its homolog remained independent. However, owing to the mechanics of meiotic divisions, the free homolog behaves as a Y chromosome (*20*). Thus, the X-autosome fusion effectively produced two Y chromosomes: the ancestral one and a "neo-Y" [formerly a Muller D autosome (*20*)]. Theory (*1*, *2*, *21*) and empirical data (*21–23*) indicate that neo-Y chromosomes degenerate, and the observation of only one Y chromosome in *D. pseudoobscura* had earlier led to the conclusion that the neo-Y was lost or fused to the Y (*20*) and that the current Y is the ancestral one. However, as we show here, the ancestral Y is now part of an autosome in *D. pseudoobscura*, leaving unresolved the origin of the current Y. An intriguing possibility is that the current Y is a degenerated neo-Y (Fig. 3). This hypothesis predicts that 3L-derived genes will be found on the current *D. pseudoobscura* Y, and indeed most of the genes we found are 3L-derived (Table 1). It remains to be seen whether they really trace back to the Y-A fusion or are more recent duplications. Generally, fusions between an autosome and a sex chromosome do not cause severe meiotic problems (*20*), but a detailed examination of one such case (*D. miranda*) revealed a low frequency (1 to 3%) of aneuploid gametes (*24*). We suggest that a second fusion, involving the other sex chromosome, could ameliorate the meiotic problems; this would explain the Y-autosome "fusion" reported here. Such a pair of fusions has occurred in *D. albomicans* (*25*); however, in that case the original Y chromosome retained its Y-chromosome status, whereas in *D. pseudoobscura* the formerly Y-linked genes became autosomal, and the current Y shares no homology with the ancestral one.

Y chromosomes are known to accumulate autosomal male-related genes both in *Drosophila* and in mammals (*3–5*). The finding that the ancestral *Drosophila* Y chromosome is now part of an autosome in the *D. pseudoobscura* lineage is thus unexpected, because the movement was in the opposite direction. The *D. pseudoobscura* Y has all the features of a typical *Drosophila* Y chromosome (including X pairing and being essential for male fertility), and yet it is less than 18 My old. In addition to the origin of Y chromosomes, the present findings have a direct bearing on the evolution of heterochromatin and intron sizes, because heterochromatin expanded in introns and intergenic regions of Y-linked genes and contracted when those Y-linked genes moved back to autosomes.

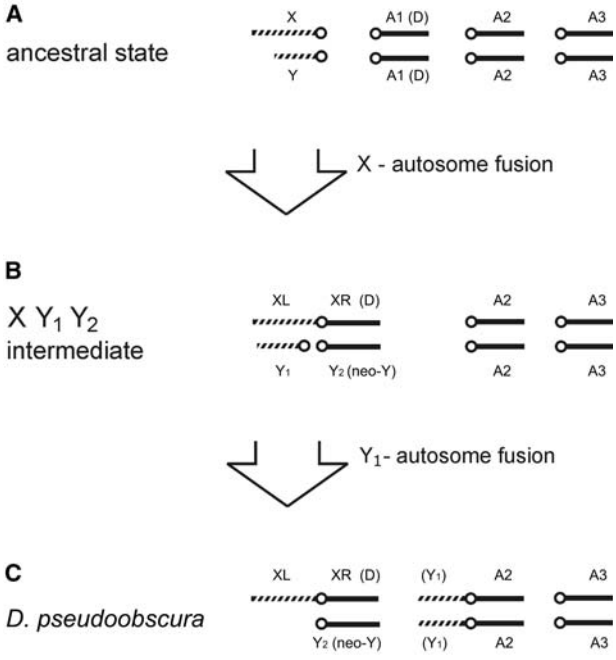


Fig. 3. A proposed model for the origin of *D. pseudoobscura* sex chromosomes. Autosomes are shown as solid bars, sex chromosomes as striped bars, and centromeres as open circles. (**A**) Ancestral state (*D. melanogaster* and obscura subgroup species), with autosomal pairs (shown three: A1, A2, and A3) and the sex chromosomes. A centric fusion between the X and one autosome (Muller element D, corresponding to the 3L in *D. melanogaster*) generated an X Y_1 Y_2 sex-determination system, shown in (**B**). This system was transient in the *D. pseudoobscura* lineage, but exists in many species (*20*). A second fusion (or large translocation) between the ancestral Y (Y_1) and another autosome (still unidentified) led to the current state of *D. pseudoobscura* and closely related species [pseudoobscura and affinis subgroups; step (**C**)]. The order of the fusions and the origin of the current Y (shown here as the neo-Y) are hypothetical, but it is certain that formerly Y-linked genes now reside on an autosome. For the sake of simplicity we represented this event as a centric fusion, but it may instead be one or more translocation events. The proposed origin of the current Y (from Muller element D) predicts that it pairs during meiosis with the autosomally derived arm of the X (XR), rather than with the ancestral X (XL).

References and Notes

1. J. J. Bull, *Evolution of Sex Determining Mechanisms* (Benjamin Cummings, Menlo Park, CA, 1983).
2. B. Charlesworth, D. Charlesworth, *Philos. Trans. R. Soc. London B Biol. Sci.* **355**, 1563 (2000).
3. H. Skaletsky *et al.*, *Nature* **423**, 825 (2003).
4. A. B. Carvalho, B. A. Dobo, M. D. Vibranovski, A. G. Clark, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13225 (2001).
5. A. B. Carvalho, *Curr. Opin. Genet. Dev.* **12**, 664 (2002).
6. A. I. Kalmykova, Y. Y. Shevelyov, A. A. Dobritsa, V. A. Gvozdev, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6297 (1997).
7. M. Steinemann, W. Pinsker, D. Sperlich, *Chromosoma* **91**, 46 (1984).
8. M. A. Baffi, C. R. Ceron, *Biochem. Genet.* **40**, 411 (2002).
9. J. H. P. Hackstein, R. Hochstenbach, E. Hauschteck-Jungen, L. W. Beukeboom, *Bioessays* **18**, 317 (1996).
10. K. Tamura, S. Subramanian, S. Kumar, *Mol. Biol. Evol.* **21**, 36 (2004).
11. R. Kurek, A. M. Reugels, U. Lammermann, H. Bünemann, *Genetica* **109**, 113 (2000).
12. H. Bünemann, personal communication.
13. S. Richards *et al.*, *Genome Res.*, in press.
14. Materials and methods are available as supporting material on *Science* Online.
15. P. M. O'Grady, *Mol. Phylogenet. Evol.* **12**, 124 (1999).
16. K. Goddard, A. Caccone, J. R. Powell, *Evolution* **44**, 1656 (1990).
17. R. A. Voelker, K. I. Kojima, *Evolution* **25**, 119 (1971).
18. M. Gatti, S. Pimpinelli, *Chromosoma* **88**, 349 (1983).
19. A. B. Carvalho *et al.*, *Genetica* **117**, 227 (2003).
20. M. J. D. White, *Animal Cytology and Evolution* (Cambridge Univ. Press, Cambridge, UK, 1973).
21. W. R. Rice, *Bioscience* **46**, 331 (1996).
22. M. Steinemann, S. Steinemann, F. Lottspeich, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5737 (1993).
23. D. Bachtrog, *Nat. Genet.* **34**, 215 (2003).
24. K. Cooper, *Genetics* **31**, 181 (1946).
25. Y. C. Yu, F. J. Lin, H. Y. Chang, *Heredity* **83**, 39 (1999).

26. All genomic sequence was produced by the Human Genome Sequencing Center at Baylor College of Medicine (HGSC-BCM) and assembled there and at The Institute for Genomic Research (TIGR). We thank S. Richards, Y. Liu, and R. Gibbs (HGSC-BCM), and S. Salzberg, A. Delcher, and M. Pop (TIGR), for providing assembly information; X. Wang, Y. Hatta-Ohashi, F. Krsticevic, C. Lazoski, and D. R. Azevedo for technical assistance; D. Lindsley, R. Hoskins, R. Guigó, D. Bachtrog, K. Montooth, P. Wittkopp, and C. Aquadro for helpful comments; and Pennsylvania State University for computer facilities. This work was supported by grants from the NIH (GM64590 and TW005673) and Conselho Nacional de Desenvolvimento Científico e Tecnológico.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1101675/DC1
Materials and Methods
SOM Text
Figs. S1 to S5
Tables S1 and S2
References

18 June 2004; accepted 27 October 2004
Published online 4 November 2004;
10.1126/science.1101675
Include this information when citing this paper.

Nutrient-Specific Foraging in Invertebrate Predators

David Mayntz,[1,2]* David Raubenheimer,[1,3] Mor Salomon,[4] Søren Toft,[2] Stephen J. Simpson[1]†

Many herbivores and omnivores adjust their food selection behavior to regulate the intake of multiple nutrients. Carnivores, however, are generally assumed to optimize the rate of prey capture rather than select prey according to nutrient composition. We showed experimentally that invertebrate predators can forage selectively for protein and lipids to redress specific nutritional imbalances. This selection can take place at different stages of prey handling: The predator may select among foods of different nutritional composition, eat more of a prey if it is rich in nutrients that the predator is deficient in, or extract specific nutrients from a single prey item.

It is generally believed that the body composition of prey animals is nutritionally balanced for carnivores, which consequently should be limited by the availability rather than the nutritional balance of prey (*1*–*3*). Herbivores and omnivores, by contrast, are known to face a heterogeneous nutritional environment and therefore possess a well-developed ability to balance their intake of multiple nutrients (*4*, *5*). There is, however, accumulating evidence to suggest that some carnivores too might benefit from having nutrient-specific regulatory abilities. For example, different insect species vary widely in their body composition (*6*–*8*), and such variation can affect the performance of their invertebrate predators (*9*, *10*). One species of spider has been observed to mix its intake from several prey species, leading to the yet untested hypothesis that they do so to compose diets with optimal amino acid makeup (*11*). Moreover, some domesticated carnivorous fish can compose diets from pure macronutrient sources in proportions that differ from random selection (*12*, *13*), but these fish were unable to compensate for enforced nutritional imbalance by differentially selecting among foods (*12*).

We tested whether invertebrate predators can forage selectively for nutrients in order to redress specific nutritional imbalances. We investigated three polyphagous predators that use very different hunting tactics for catching prey: a highly mobile ground beetle, an ambushing wolf spider of intermediate mobility, and a web-building spider. The general approach was, first, to manipulate the nutritional state of the predators by feeding them a pretreatment diet with either a high or low ratio of protein to lipid for 1 or 2 days. Thereafter, we tested their feeding responses to the pretreatment food and/or a nutritionally complementary alternative.

In the case of the ground beetle *Agonum dorsale* (Coleoptera: Carabidae), we offered both the pretreatment food and the alternative food simultaneously, to emulate the scenario of a roving predator that actively selects between different prey types. In this experiment, semisynthetic food (based on powdered locusts) was used to avoid treatment-specific predator avoidance behavior that might result from the use of live prey. The results showed that intake of the two test foods was strongly dependent on the nutrient composition of the pretreatment diet (Fig. 1). Beetles pre-fed a lipid-enriched diet subsequently consumed more protein-rich than lipid-rich food, and those pre-fed with a protein-enriched diet preferentially consumed food with high lipid content, as well as eating more food overall. Thus, the beetles were able to select among foods of different pro-

[1]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. [2]Department of Ecology and Genetics, Aarhus University, 8000 Århus C, Denmark. [3]School of Biological Sciences and Psychology Department, University of Auckland, New Zealand, Private Bag 92019, Auckland, New Zealand. [4]Life Sciences Department and Mitrani Department of Desert Ecology, Ben-Gurion University of the Negev, Israel.

*To whom correspondence should be addressed. E-mail: david.mayntz@zoology.oxford.ac.uk
†Present address: School of Biological Sciences, University of Sydney, Heydon-Laurence Building, A08, Sydney, New South Wales 2006, Australia.

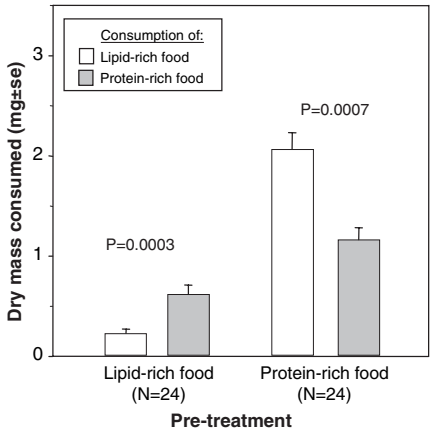


Fig. 1. Nutrient balancing by the predatory ground beetle *A. dorsale*, when given a free choice between lipid-rich or protein-rich foods. Beetles were pre-fed with one of the two food types for 48 hours and were subsequently tested on both foods at the same time for 48 hours. The two food types were dried and crushed locusts mixed with either protein (protein-rich food) or lipids (lipid-rich food) (*15*). Values shown are means + SEM; *P* values test for equal consumption of lipid-rich and protein-rich foods (paired *t* tests).

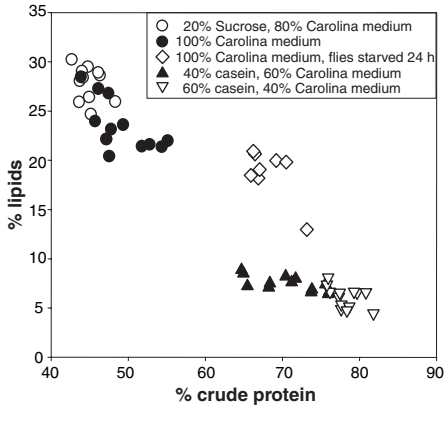


Fig. 2. Relative composition of lipid and crude protein in adult *Drosophila melanogaster* fruit flies. Variations in body compositions were produced by varying the nutrient composition of the fruit fly cultures or by starving flies for 24 hours at 20°C. Fruit fly media were produced from a basis of Carolina medium, to which additional casein or sucrose was added (*15*). Each point represents a sample of 100 ± 5 male and female flies from a single culture bottle.

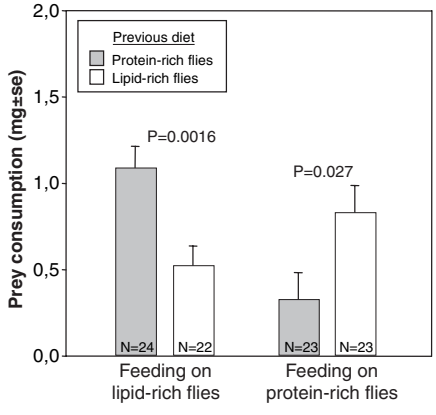


Fig. 3. Intake of protein-rich or lipid-rich flies (dry mass) over 72 hours in no-choice tests by wolf spiders, *P. prativaga*. Before testing, spiders were pre-fed one of the same two prey types for 24 hours. Protein-rich flies were reared on a mixture of 40% Carolina medium and 60% casein; lipid-rich flies were reared on pure Carolina medium (*15*). Values shown are means + SEM; *P* values test for equal consumption by spiders with different feeding history (*t* tests).

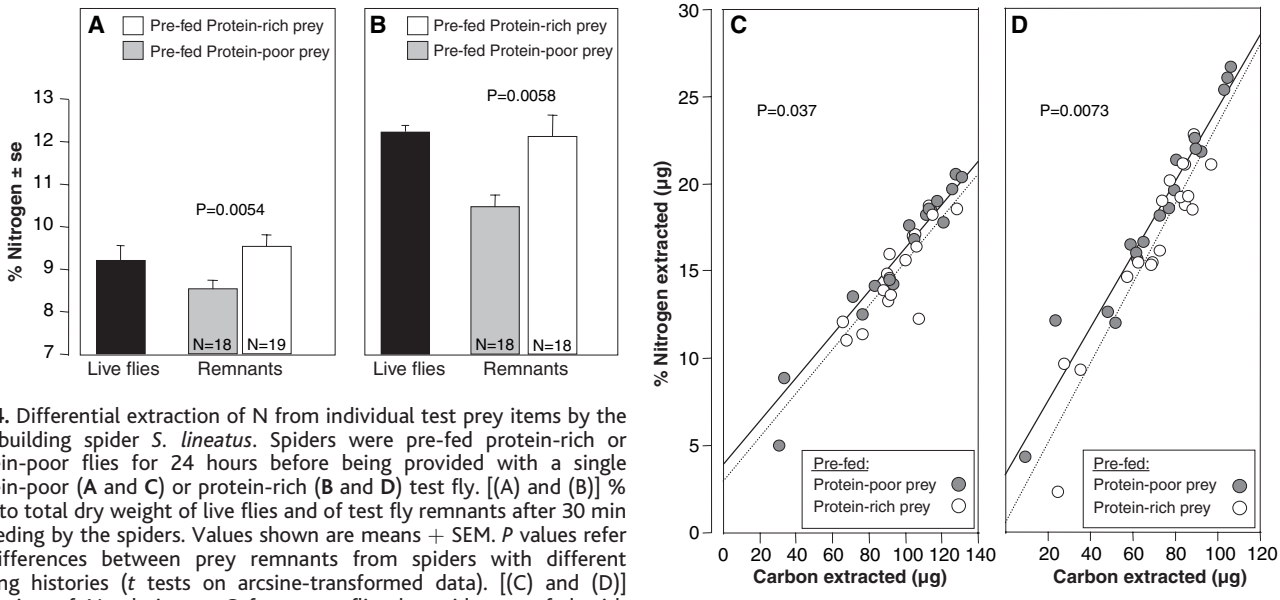


Fig. 4. Differential extraction of N from individual test prey items by the web-building spider *S. lineatus*. Spiders were pre-fed protein-rich or protein-poor flies for 24 hours before being provided with a single protein-poor (**A** and **C**) or protein-rich (**B** and **D**) test fly. [(A) and (B)] % of N to total dry weight of live flies and of test fly remnants after 30 min of feeding by the spiders. Values shown are means + SEM. *P* values refer to differences between prey remnants from spiders with different feeding histories (*t* tests on arcsine-transformed data). [(C) and (D)] Extraction of N relative to C from test flies by spiders pre-fed with protein-rich (dashed lines) or protein-poor (solid lines) flies. The regression lines were not significantly different from parallel, either when test prey was protein-poor (C) or when it was protein-rich (D) [$P > 0.05$, analysis of covariance (ANCOVA) test for homogeneity of slopes]. *P* values test for different elevations of regression lines (ANCOVA, test for different intercepts) by spiders with different feeding histories. Different elevation indicates differences in the extraction of N relative to C (*16*).

tein and lipid composition to compensate for previous imbalances of these nutrients.

The wolf spider *Pardosa prativaga* (Araneae: Lycosidae) is a sit-and-wait predator that frequently moves between foraging patches. To emulate a scenario where foraging patches vary in the quality of available food, we offered these predators only one prey type (either the pretreatment or the complementary prey type) and recorded how much of each was eaten. Food was live *Drosophila* fruit flies that had been raised on different media, resulting in body compositions with either a high or low protein-to-lipid ratio (Fig. 2). The data show that wolf spiders consumed more prey mass if the prey contained a high concentration of the nutrient that was in low concentration in their previous meal (Fig. 3). When feeding on lipid-rich flies, wolf spiders ate more if their previous diet was protein-rich flies; but when test prey were protein-rich, the wolf spiders ate more if their previous diet was lipid-rich. Like the beetles, these predators redressed a nutrient-specific imbalance but did so by varying their consumption of only a single prey type.

The desert spider *Stegodyphus lineatus* (Araneae: Eresidae) builds durable webs for prey capture and spends relatively long periods at a single hunting site. Prey items arrive sporadically, and their nutritional composition is beyond the control of the spider. We therefore tested whether these spiders were able to extract nutrients differentially from within a single prey item to redress a nutritional imbalance. The spiders' nutritional state was manipulated by feeding them either lipid-rich or protein-rich flies for 24 hours. We subsequently measured how efficiently the spiders extracted nitrogen (N) and carbon (C) from a single prey item during 30 min of feeding. We found that the percentage of N left in the fly remnants after feeding was lower if the spiders' previous prey was protein-poor than if it was protein-rich, and this effect appeared when spiders were tested on both protein-poor (Fig. 4A) and protein-rich (Fig. 4B) flies. Furthermore, the extent to which spiders extracted N relative to C (N extracted per unit of C extracted) varied with feeding history: More N was extracted per unit of C if the spiders were previously fed protein-poor prey, both when test prey were protein-poor (Fig. 4C) and when they were protein-rich (Fig. 4D). This shows that these spiders have the ability to compensate for nutritional imbalances by extracting nutrients selectively from within individual prey items.

Our study shows that invertebrate predators can regulate their intake of protein and lipids, and they can use this ability to redress an existing imbalance of these nutrients. The fact that only relatively short periods of prefeeding (24 to 48 hours) were sufficient to generate nutrient-specific compensatory responses suggests that the observed regulatory mechanisms are involved in fine-tuned compensation for day-to-day variations in prey nutrient composition. The need for such a capacity in insectivorous predators is highlighted by the demonstration that prey composition varies markedly with rearing diet, even within a single species (Fig. 2). Our data show, furthermore, that nutrient selection may occur at several stages of prey handling: either by selecting among foods of different nutritional composition (ground beetles), by adjusting consumption of a single prey type depending on its nutrient composition (wolf spiders), or by extracting nutrients selectively from within individual prey items (web-builders). These results are highly relevant for optimal foraging models, where it is generally assumed that for carnivores, prey quantity rather than quality matters (*2*). They are also relevant to ecological stoichiometry, in which models usually assume that prey animals (such as the flies in our study) maintain a constant body composition in the face of variable diets (*14*).

References and Notes

1. B. G. Galef, *Neurosci. Biobehav. Rev.* **20**, 67 (1996).
2. D. W. Stephens, J. R. Krebs, *Foraging Theory* (monographs in behavior and ecology, Princeton Univ. Press, Princeton, NJ, 1986).
3. S. Slansky, J. M. Scriber, in *Comprehensive Insect Physiology, Biochemistry and Pharmacology*, vol. 4, *Regulation: Digestion, Nutrition, Excretion*, G. A. Kerkut, L. I. Gilbert, Eds. (Pergamon, Oxford, 1985), pp. 87–163.
4. D. Raubenheimer, S. J. Simpson, *Nutr. Res. Rev.* **10**, 151 (1997).
5. H. R. Berthoud, R. J. Seely, *Neural and Metabolic Control of Macronutrient Intake* (CRC Press, Boca Raton, FL, 2000).
6. W. F. Fagan *et al.*, *Am. Nat.* **160**, 784 (2002).
7. D. Barker, M. P. Fitzpatrick, E. S. Dierenfeld, *Zoo Biol.* **17**, 123 (1998).
8. K. H. Redford, J. G. Dorea, *J. Zool.* **203**, 385 (1984).
9. S. Toft, *J. Arachnol.* **27**, 301 (1999).
10. D. Mayntz, S. Toft, *Oecologia* **127**, 207 (2001).
11. M. H. Greenstone, *Nature* **282**, 501 (1979).
12. M. Vivas, F. J. Sanchez-Vazquez, B. G. Garcia, J. A. Madrid, *Aquac. Res.* **34**, 271 (2003).
13. F. J. Sanchez-Vazquez, T. Yamamoto, T. Akiyama, J. A. Madrid, M. Tabata, *Physiol. Behav.* **66**, 45 (1999).

14. R. W. Sterner, J. J. Elser, *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere* (Princeton Univ. Press, Princeton, NJ, 2002).
15. Materials and methods are available as supporting material on *Science* Online.
16. D. Raubenheimer, S. J. Simpson, *Entomol. Exp. Appl.* **62**, 221 (1992).
17. We thank E. Rasmussen for rearing prey animals and P. Henriksen for conducting elemental analyses. This work was supported by a grant from the Carlsberg Foundation to D.M., from the Danish Natural Science Research Council to S.T., from the British Biotechnology and Biological Sciences Research Council to S.J.S. and D.R., and by a Danish Government Scholarship for Foreign Nationals to M.S.. The first draft of this paper was written while D. Raubenheimer and D. Mayntz were fellows at the Center of Advanced Studies in Oslo.

Supporting Online Material
www.sciencemag.org/cgi/content/full/307/5706/111/DC1
Materials and Methods
Reference

21 September 2004; accepted 8 November 2004
10.1126/science.1105493

Disulfide Isomerization After Membrane Release of Its SAR Domain Activates P1 Lysozyme

Min Xu,* Arockiasamy Arulandu,* Douglas K. Struck, Stephanie Swanson, James C. Sacchettini,† Ry Young†

The P1 lysozyme Lyz is secreted to the periplasm of *Escherichia coli* and accumulates in an inactive membrane-tethered form. Genetic and biochemical experiments show that, when released from the bilayer, Lyz is activated by an intramolecular thiol-disulfide isomerization, which requires a cysteine in its N-terminal SAR (signal-arrest-release) domain. Crystal structures confirm the alternative disulfide linkages in the two forms of Lyz and reveal dramatic conformational differences in the catalytic domain. Thus, the exported P1 endolysin is kept inactive by three levels of control—topological, conformational, and covalent—until its release from the membrane is triggered by the P1 holin.

In phage infections, the timing of host lysis is strictly regulated (*1*, *2*). For double-stranded DNA phages, lysis requires degradation of the cell wall by a phage-encoded endolysin, or lysozyme. In the classic systems of phages λ and T4, premature cell wall degradation is prevented by keeping the bacteriolytic enzyme sequestered in the cytoplasm (*3*). At a genetically programmed time, the holin, a small phage-encoded membrane protein, disrupts the membrane, which allows access of the lysozyme to the cell wall; lysis follows within seconds (*1*, *4*). However, endolysins are not always dependent on holins for export (*5*–*7*). The endolysins from enterobacteriophage P1 (Lyz) and lambdoid coliphage 21 (R^{21}), although orthologs of T4 lysozyme (*8*, *9*), are exported by the host *sec* system by virtue of an N-terminal transmembrane domain (TMD) (Fig. 1) that functions as a type II signal anchor, or uncleaved signal peptide (*7*, *10*). This leaves the secreted protein in the periplasm but tethered to the membrane in an inactive form. When the membrane is depolarized by the holin, this N-terminal TMD exits the bilayer, resulting in activation of the endolysin and cell lysis. Because of the ability to escape from the membrane, the Lyz and R^{21} TMDs, as well as similar motifs in many other phage lysozymes, including the well-studied coliphages Mu (*7*) and T1 (*11*), have been designated as SAR ("signal arrest and release") domains. The capacity for membrane release probably derives from a high content of the small hydrophobic and polar residues Gly, Ala, Ser, and Thr, but the mechanism for extraction from the bilayer is unknown.

It was also not understood how the tethered SAR endolysins are kept inactive until holin triggering. To address this issue, we aligned Lyz with T4 lysozyme, gp*e* (Fig. 1). Members of the T4 lysozyme family have an essential catalytic triad corresponding to Glu^{11}, Asp^{20}, and Thr^{26} of gp*e* (*12*). Although Lyz has two of these conserved catalytic residues, Glu^{42} and Thr^{57}, the alignment suggests that Cys^{51} in Lyz fulfills the catalytic role of Asp^{20} in T4 lysozyme. Consistent with this, the gp*e*D20C (*13*) variant is functional (*14*). The functionality of LyzC51D and the sensitivity of Lyz, but not LyzC51D, to thiol reagents confirmed the catalytic role of Cys^{51} (Fig. 2A and table S1). In addition to the catalytic Cys^{51}, Lyz has six other cysteine residues, one of which, Cys^{13}, is located in the N-terminal SAR domain (Fig. 1) and is expected to be embedded in the bilayer in its sulfhydryl form. To test this, chemical cleavage was used to determine the positions of reduced cysteines in Lyz present in induced cultures before and after the onset of lysis (*15*). Reduced cysteines, but not cysteines in disulfide bonds, can be cyanylated, and the cyanylated cysteines can be sites of partial cleavage when treated with ammonium hydroxide (*16*). In samples taken soon after induction, only a single cleavage product was detected, with an apparent mass expected from cleavage at Cys^{13} (Fig. 3A). This indicates that, in the membrane-associated, inactive form, Cys^{13} is present in its sulfhydryl form, but all of the six periplasmic cysteines, including the catalytic Cys^{51}, are in disulfides. In samples taken after the onset of lysis, two cleavage products were detected, one generated by cleavage only at Cys^{13}, the other by cleavage only at Cys^{51}. Peptides corresponding to cleavage at the other five cysteine residues were never detected. Because none of the cysteines in Lyz are involved in intermolecular disulfide bonds (fig. S1), cleavage at Cys^{13} and Cys^{51} in the post lysis sample indicates the presence of two isomeric forms of Lyz differing in the arrangement of intramolecular disulfide bonds. By this interpretation, the membrane-associated form of Lyz would have a single free cysteine, Cys^{13}, in the membrane-embedded SAR domain. By contrast, the soluble and active form of Lyz would have a sulfhydryl at Cys^{51} whereas all other cysteines, including Cys^{13}, would exist in disulfide linkages.

These results suggest a model for Lyz activation involving disulfide bond isomerization with no net change in overall oxidation state and in which a critical feature is the liberation of the Cys^{13} sulfhydryl from the bilayer (Fig. 3B). Accordingly, although mutants of Lyz in which Cys^{13} is replaced by Ser or Ala (C13S and C13A, respectively) are still released from the cytoplasmic membrane (Fig. 3A), they are lytically inactive. Moreover, the addition of low concentrations of dithiothreitol (DTT) to cultures expressing the C13S mutant causes lysis (Fig. 2B). Here, DTT probably acts by substituting for Cys^{13} in the disruption of the inactivating disulfide. The lysis defect of C13S could also be suppressed intragenically by a second change at Cys^{44}, C44S; the C13S,44S double mutant of Lyz was not only lytically active but it caused host lysis with kinetics that were slightly faster than seen with wild-type Lyz (Fig. 2B), presum-

Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843–2128, USA.

*These authors contributed equally to the results described in this work.
†To whom correspondence should be addressed. E-mail: sacchett@tamu.edu (J.C.S.) and ryland@tamu.edu (R.Y.).

ably because one step in the activation, disulfide bond isomerization, is unnecessary. This implicates Cys[44] as the cysteine that forms a disulfide with Cys[51] in the inactive, nascent form of the endolysin. The Lyz C44S single mutant is inactive (Fig. 2B), presumably because, lacking Cys[44], its normal disulfide partner, Cys[51] is involved in incorrect disulfide bonds before and after membrane release.

Further support for this model of Lyz activation is derived from the behavior of the chimera R[21]ΦLyz, in which the SAR domain of Lyz is replaced by that of R[21], the SAR-type lysozyme of phage 21 [(Fig. 1); also, see (*7*)]. The SAR domain of R[21] has no significant sequence similarity to the Lyz SAR and does not contain cysteine, presumably because R[21] has the canonical Glu-Asp-Thr catalytic triad and thus lacks a catalytic Cys residue (Fig. 1). The R[21] SAR domain supports export and release of the R[21]ΦLyz chimera (*7*), but the released protein is inactive. The activity could be restored by the introduction of C44S into the chimera (Fig. 2C) or the addition of reducing agent to the induced culture (*17*), which indicates that the inhibitory Cys[44] to Cys[51] disulfide bond is retained when the chimera is released from the membrane. Moreover, introduction of a Cys into the R[21] SAR domain at position 13 (but not position 16) also restores the lytic activity of the chimera (Fig. 2C). Hence, on release from the bilayer, a heterologous SAR domain can provide the thiol required for reduction of the Cys[44] to Cys[51] disulfide bond. The lack of a strict sequence context for the activating thiol was also observed in intragenic suppressors of the Lyz C13S mutant, where substitution of a cysteine at positions 9 to 11, 14, or 15 (but not 12, 16, or 17) resulted in enzymatically active double mutants (Fig. 2, D and E). Although there is some steric constraint in the position of the activating sulfhydryl group, these results suggest that there is little requirement for specific contacts between the SAR domain, after its extraction from the bilayer, and the periplasmic domain of Lyz.

These results also imply that, other than providing an export signal and supplying the thiol necessary for reduction of the inhibitory disulfide bond, the Lyz SAR domain is not essential to the structure of the active endolysin. To test this idea, we replaced the entire N-terminal segment of Lyz, including the SAR domain, with the cleavable signal sequence of a well-characterized secreted protein, alkaline phosphatase (PhoA) (*18*) (Fig. 1). When this construct was expressed in *E. coli*, the PhoAΦLyz protein was secreted to the periplasm with its signal peptide removed (Fig. 3A), but lysis required the addition of DTT (Fig. 2F). Note that the C[51] replaced by Asp (C51D) allele of *phoAΦlyz* caused host lysis even in the absence of a reductant. Hence, the SAR domain of Lyz is not essential for the function of the mature form of Lyz.

To dissect the structural basis of Lyz activation, we determined the crystal structures of the full-length, soluble, active endolysin, Lyz[a], and also of the inactive, secreted, and processed form of PhoAΦLyz, Lyz[i], which consists of residues 29 to 185. Crystals of Lyz[a] and Lyz[i] were obtained in different space groups (table S2). In the Lyz[a]

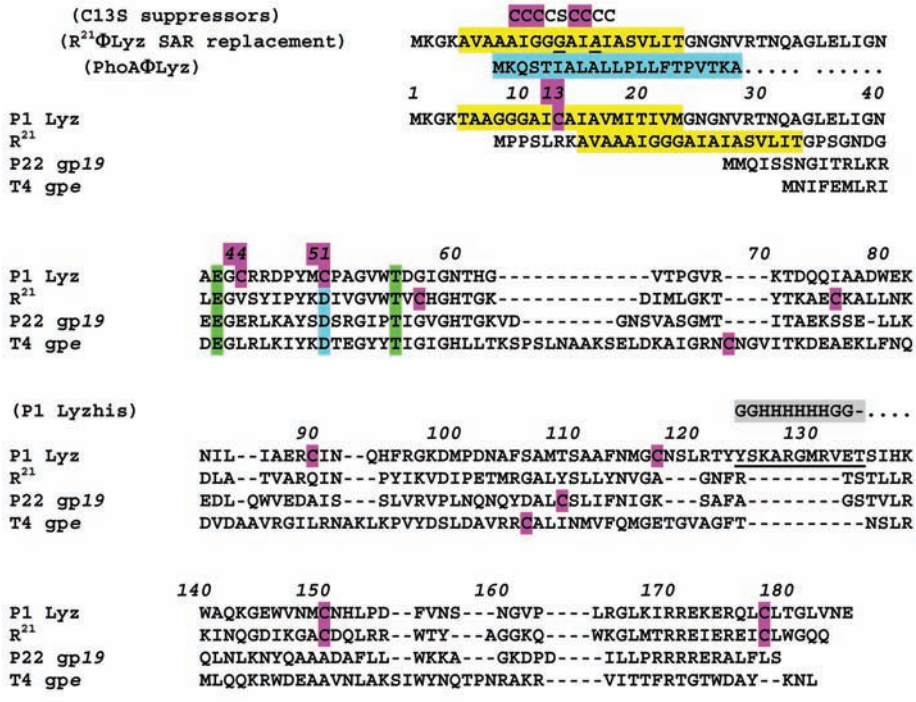


Fig. 1. SAR endolysins. P1 Lyz and R[21], the endolysin from lambdoid phage 21, are aligned with the homologous soluble endolysins T4 gpe and P22 gp*19* (*15*). Residue numbering (italics) is for P1 Lyz. Coloring: purple, Cys; green, catalytic Glu, Thr; blue, catalytic Asp; yellow: predicted SAR TMDs. The positions of Cys substitutions made in the C13S allele are indicated; purple represents changes that suppressed the lysis defect. Below the Cys substitutions, the R[21]ΦLyz chimera is shown. Positions where a Cys substitution suppresses (Gly[13]) or fails to suppress (Ala[16]) the lysis defect of the chimera are underlined and underlined italic, respectively. The signal sequence of PhoA is shown in blue at the position where it replaces the N terminus of P1 Lyz, in the PhoAΦLyz chimera. The $G_2H_6G_2$ shaded in gray in LyzHis replaces residues 124 to 134 (underlined) of Lyz.

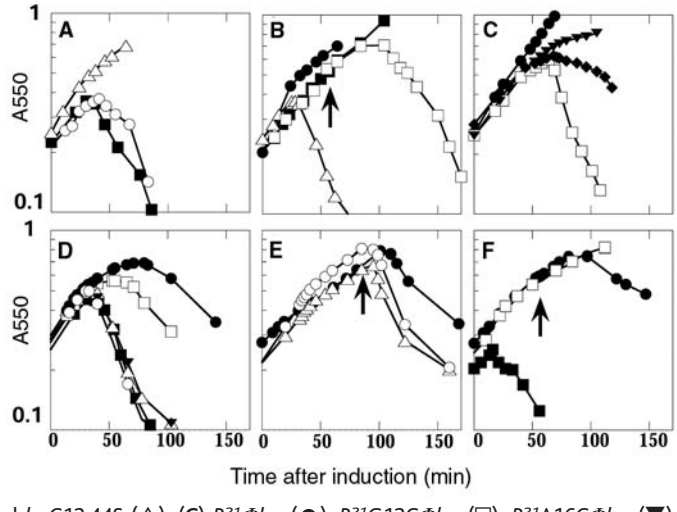


Fig. 2. The roles of Cys[13], Cys[44], and Cys[51] in Lyz function. XL1-Blue cells carrying plasmids with the indicated *lyz* allele were induced with IPTG and monitored for turbidity. All plasmids were isogenic to pJFLyz (*lyz*$^+$). Residue numbering refers to positions in Lyz unless otherwise indicated. Vertical arrows indicate time of addition of 1 mM DTT. (**A**) *lyz*$^+$ (■), *lyz*C51S (△), and *lyz*C51D (○). (**B**) *lyz*C13S (■); *lyz*C13S with the addition of DTT (□); *lyz*C44S (●); and *lyz*C13,44S (△). (**C**) *R[21]Φlyz* (●), *R[21]*G13C*Φlyz* (□), *R[21]*A16C*Φlyz* (▼), and *R[21]Φlyz*C44S (◆). (Positions here refer to the position in the chimera, not R[21]; see Fig. 1.) (**D**) *lyz*$^+$ (■); *lyz*C13S,G9C (□); *lyz*C13S,G10C (▼); *lyz*C13S,A11C (●); *lyz*C13S,A14C (△); and *lyz*C13S,I15C (○). (**E**) *lyz*C13S,I12C (●); *lyz*C13S,A16C (□); and *lyz*C13S,V17C (△). (**F**) *phoAΦlyz* (□); *phoAΦlyz*, with the addition of DTT (●); and *phoAΦlyz*C51D(■).

structure (Fig. 4A), which was solved using multiwavelength anomalous diffraction (MAD) methods (*19*) to 1.8 Å resolution, the entire protein is visible except residues 1 to 8. With the exception of the N-terminal SAR domain and two β hairpins protruding from the C-terminal domain (Fig. 4A; fig. S2; SOM text), the structure of Lyz^a is very similar to T4 gp*e* lysozyme, composed of two domains connected by a long interdomain helix (residues 74 to 96) (supporting online text). Part of the SAR TMD that has been released from the membrane persists in an α helix (residues 15 to 25), whereas its N-proximal residues are in an extended conformation. The crystal structure confirms that Cys^{51} of the active form of Lyz is in sulfhydryl form and exposed to solvent, occupying the site corresponding to Asp^{20} in gp*e*. As predicted, Cys^{13} of the SAR domain and Cys^{44} are in a disulfide linkage, which holds the SAR helix close to the body of the endolysin; also as predicted, there are few specific contacts between the SAR domain and the rest of Lyz^a.

The Lyz^i structure was solved by molecular replacement using the Lyz^a structure as a search model and refined to 1.1 Å (Fig. 4B). The two structures are very similar for residues 29 to 39 and the entire C-terminal domain (residues 88 to 185), with a root mean square deviation (rmsd) of 0.65 Å. However, the bulk of the N-terminal domain (residues 40 to 73) and part of the interdomain connecting helix (residues 74 to 85) are very different in the two structures (rmsd for residues 40 to 85 = 7.62 Å) (fig. S3). The Lyz^i structure confirms that Cys^{44} and Cys^{51} are in a disulfide linkage, which shows clear electron density in the refined structure. This disulfide bond holds the N-terminal domain in a compact, closed conformation in which the entire active site cleft is absent, and the distinct separation of the N- and C-terminal domains is missing (Fig. 4, C and D; fig. S4). In Lyz^i, two of the catalytic residues, Glu^{42} and Thr^{57}, are on the surface of the protein with the side chains facing outwards (Fig. 4, B and C).

On the basis of these structures, it is clear that major movements within the region Gly^{39} to Asn^{84} are required for the transition from the inactive to active forms (Fig. 4, C and D; figs. S3 and S5). The transition is punctuated by a rotation of 112° between Gly^{39} and Asn^{40} that adds another turn at the end of α2; causes α helices 3, 4, and 5 to unwind; and allows three β strands to form a β sheet (Fig. 4, C and D; fig. S5). Also the apparently strained 60° bend at Asn^{84} in the interdomain connecting helix, α6, of Lyz^i is straightened in Lyz^a. Importantly, these changes do not just unblock the active site by moving or removal of rigid domains, as in zymogen activation (*20*). Instead, formation of the catalytic cleft is accomplished by remodeling discrete secondary structure elements of the inactive form (Fig. 4, C and D; figs. S5 and S6; movie S1).

The SAR domain of Lyz has two functions. First, it mediates association of Lyz with the cytoplasmic membrane by acting as a secretory signal and a TMD. Second, when released from the membrane, its cysteine residue participates in an isomerization event necessary for Lyz to assume its active conformation. Although changes in the oxidation state of cysteine residues are known to regulate the activity of numerous proteins (*21*, *22*), the distinguishing feature here is that Lyz activation is redox-independent and proceeds without a change in its overall oxidation state.

It remains to be determined whether host proteins are involved in the membrane release, conformational changes, and disulfide bond isomerization required for the maturation of Lyz. In *E. coli*, periplasmic disulfide bond isomerases (PDI) DsbC and DsbG function by using free sulfhydryls to attack disulfide bonds in nascent or misfolded substrate proteins (*23*). Lyz-mediated lysis requires neither DsbC or DsbG (*24*), which suggests that, once liberated from the membrane, Cys^{13} may act in cis analogously to the thiol of a PDI. However, in the structure of the periplasmic domain of inactive Lyz, the Cys^{44} to Cys^{51} disulfide linkage is so buried that it would seem to be inaccessible to attack by Cys^{13}. The conformational changes required to expose this disulfide to attack and for the subsequent formation of active Lyz may be facilitated by periplasmic foldases or chaperones. A genetic selection for host mutants insensitive to the lethal lytic effect of Lyz could reveal which, if any, of these factors are involved.

Lyz is activated by relief of topological, covalent, and conformational constraints. Although many other SAR lysozymes appear to share this regulatory scheme, with a Cys residue in the predicted SAR domain and a

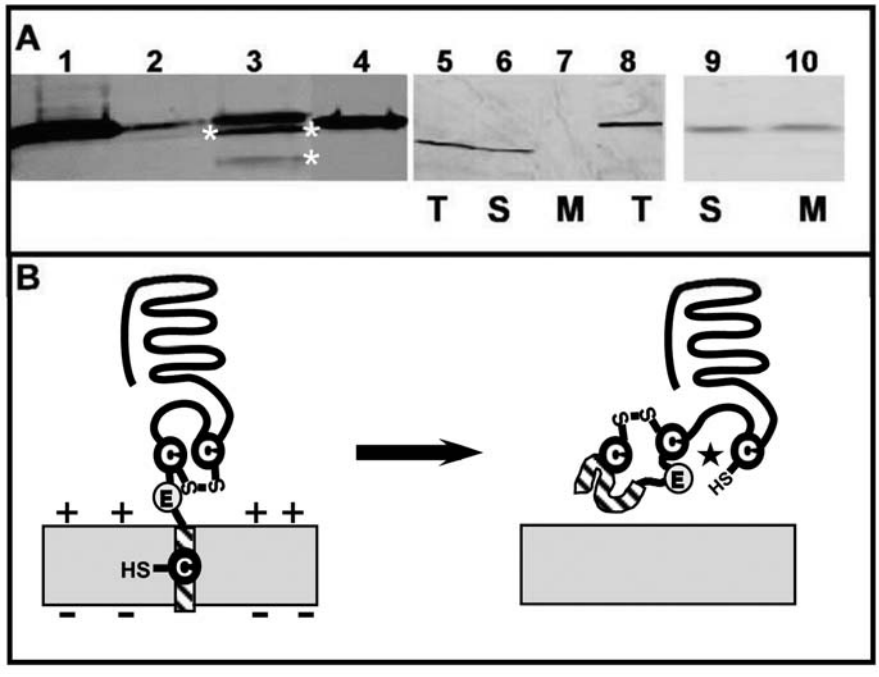


Fig. 3. (**A**) The free thiols and subcellular localization of Lyz. Lanes 1 to 4, thiol cleavage. Cells harboring pJFLyzHis were induced and samples were taken from cells carrying pJFLyzHis at 30 min (before lysis) and 80 min (after lysis) after induction, precipitated with trichloroacetic acid, subjected to cyanylation and NH_4OH cleavage, and analyzed by SDS–polyacrylamide electrophoresis (SDS-PAGE) (*15*). Lane 1, untreated Lyz; lanes 2 and 3, cleavage products from 30-min and 80-min samples, respectively; lane 4, same as lane 3 except that the cyanylation was reversed by treatment with DTT before exposure to 1 M NH_4OH. Asterisks adjacent to lanes 2 and 3 indicate bands corresponding to cleavages at Cys^{13} (top) and Cys^{51} (bottom). Lanes 5 to 10, subcellular fractionations. Total protein (T: lanes 5 and 8), soluble protein (S: lanes 6 and 9), and membrane protein fractions (M: lanes 7, 10) were prepared from induced culture of XL1-Blue cells carrying pPhoAϕLyz (lanes 5 to 7), pLyz (lane 8), or pLyzC13S (lanes 9 to 10), and analyzed by SDS-PAGE and immunoblotting (*15*). In all cases, the M_r of the major band was identical, except for lanes 5 and 6, where the signal sequence is cleaved from the PhoAΦLyz chimera. (**B**) Model for intramolecular disulfide bond isomerization after the release of the SAR domain from the membrane. Left, Lyz with its SAR TMD embedded in the fully energized membrane. Cys^{13} is in the sulfhydryl state, whereas Cys^{44} and Cys^{51} are in a disulfide linkage that prevents formation of the active site cleft. Right, Lyz after the SAR domain has been released from the depolarized membrane, undergoes refolding and disulfide bond isomerization, forming the active site (star).

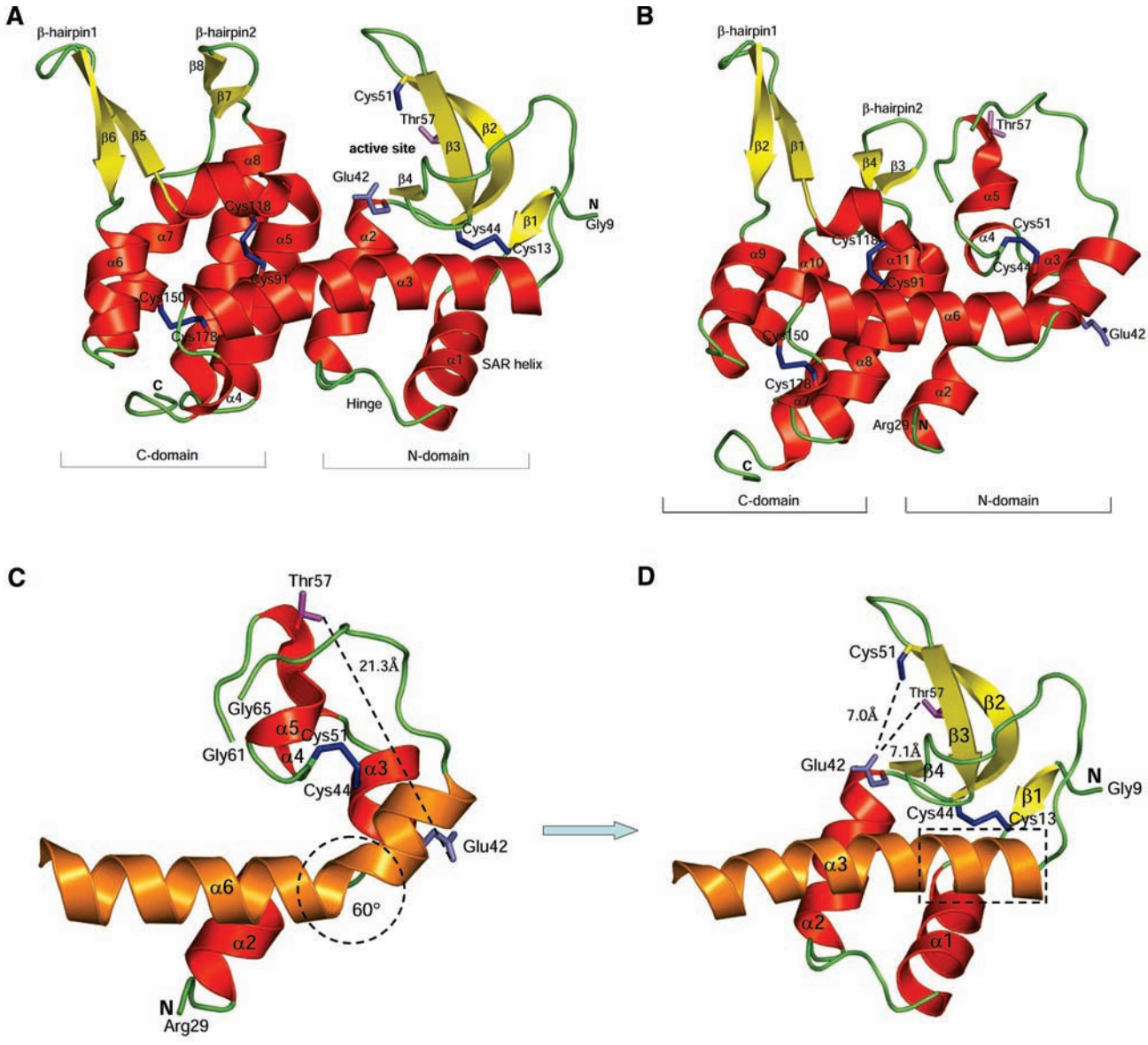


Fig. 4. Crystal structures of Lyz^a and Lyz^i: domain organization (**A** and **B**) and conformational change (**C** and **D**). (A) Structure of Lyz^a with distinct structural domains. Colors: α helices, red; β strands, yellow; loops, green; disulfides, blue. The N- and C-terminal domains are connected by an interdomain connecting helix, α3. (B) Structure of Lyz^i. Colors as in (A). The residues 1 to 28 are truncated. Lyz^i has three additional helices (α3 to α5), lacks strands β1 to β4 of Lyz^a, and has a different N-proximal disulfide (44 to 51 instead of 13 to 44). (C and D) Only the N domain and α3 [colored orange; other colors as in (A and B)] are shown. The catalytic residues Glu^{42} (slate), Cys^{51} (blue), and Thr^{57} (pink) are shown with side chains. (C) The ~60° bend in the middle of α3 is marked with a circle on Lyz^i. The catalytic Cys^{51} is in disulfide linkage with Cys^{44}. Electron density for the residues 62 to 64 is missing. (D) Creation of the active site by N-terminal domain movements and disulfide isomerization. The ~60° bend relieved in α3 is marked by a dotted box. The new disulfide (Cys^{13} to Cys^{44}) and the rearrangement of catalytic residues at the active site cleft are visible. A different view of (C and D) shows changes in α2 due to the 112° rotation at the end of that helix (fig. S6).

catalytic Cys equivalent to Lyz Cys^{51}, others, like R^{21}, do not (*7*) (fig. S7). In R^{21} and other SAR lysozymes lacking the disulfide bond regulation, the Glu residue of the catalytic triad is much closer to the predicted boundary of the SAR TMD, which suggests that in these lysozymes, membrane proximity may sterically block the active site. Moreover, in R^{21} there are 10 fewer residues between the putative membrane boundary and the active site, which suggests that the SAR domain, once extracted from the membrane, may have to provide much of or the entire α2 element in the active form and thus may interact more intimately with the rest of the catalytic domain. In any case, the redundancy and diversity of the regulatory strategies for the SAR lysozymes reflect how critical lysis timing is in the phage infection cycle.

References and Notes

1. I.-N. Wang, D. L. Smith, R. Young, *Annu. Rev. Microbiol.* **54**, 799 (2000).
2. R. Young, I.-N. Wang, W. D. Roof, *Trends Microbiol.* **8**, 120 (2000).
3. R. Young, *Microbiol. Rev.* **56**, 430 (1992).
4. A. Gründling, M. D. Manson, R. Young, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9348 (2001).
5. C. São-José, R. Parreira, G. Vieira, M. A. Santos, *J. Bacteriol.* **182**, 5823 (2000).
6. C. São-José, R. Parreira, M. A. Santos, *Recent Res. Dev. Bacteriol.* **1**, 103 (2003).
7. M. Xu, D. K. Struck, J. F. Deaton, I.-N. Wang, R. Young, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6415 (2004).
8. C. Schmidt, M. Velleman, W. Arber, *J. Bacteriol.* **178**, 1099 (1996).
9. M. T. Bonovich, R. Young, *J. Bacteriol.* **173**, 2897 (1991).
10. S. High, V. Laird, *Trends Cell Biol.* **7**, 206 (1997).
11. M. D. Roberts, N. L. Martin, A. M. Kropinski, *Virology* **318**, 245 (2004).
12. R. Kuroki, L. H. Weaver, B. W. Matthews, *Science* **262**, 2030 (1993).
13. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P,

Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
14. L. W. Hardy, A. R. Poteete, *Biochemistry* **30**, 9457 (1991).
15. Materials and methods are available as supporting material on *Science* Online.
16. J. Wu, J. T. Watson, *Protein Sci.* **6**, 391 (1997).
17. M. Xu and D. K. Struck, unpublished observations.
18. S. Michaelis, H. Inouye, D. Oliver, J. Beckwith, *J. Bacteriol.* **154**, 366 (1983).
19. W. A. Hendrickson, *Science* **254**, 51 (1991).
20. A. R. Khan, M. N. James, *Protein Sci.* **7**, 815 (1998).
21. P. J. Hogg, *Trends Biochem. Sci.* **28**, 210 (2003).
22. M. S. Paget, M. J. Buttner, *Annu. Rev. Genet.* **37**, 91 (2003).
23. H. Kadokura, F. Katzen, J. Beckwith, *Annu. Rev. Biochem.* **72**, 111 (2003).
24. M. Xu, D. K. Struck, R. Young, unpublished observations.
25. Data sets were collected at the Advanced Photon Source, Argonne National Laboratory, using the synchrotron beam-lines BIOCARS 14BM-C, 14ID-B, and SBC-CAT 19-ID, and we thank the beam-line staff for their assistance. Support for this work came from PHS grant GM27099 and Robert A. Welch Foundation grant A-1384 to R.F.Y., PHS grant GM62410 and a Wolfe-Welch chair from the Robert A. Welch Foundation to J.C.S., and the Program for Membrane Structure and Function, a Program of Excellence initiative of the Office of the Vice President for Research at Texas A&M University. Coordinates for both Lyz_i and Lyz_a have been deposited with the Protein Data Bank and are available with the ID codes 1XJU and 1XJT, respectively.

Supporting Online Material
www.sciencemag.org/cgi/content/full/307/5706/113/DC1
Materials and Methods
SOM Text
Figs. S1 to S7
Tables S1 and S2
References
Movie S1

13 September 2004; accepted 10 November 2004
10.1126/science.1105143

Requirement of Voltage-Gated Calcium Channel β_4 Subunit for T Lymphocyte Functions

Abdallah Badou,[1] Srisaila Basavappa,[2,4] Rooma Desai,[3] You-Qing Peng,[4] Didi Matza,[1] Wajahat Z. Mehal,[1] Leonard K. Kaczmarek,[2,3] Emile L. Boulpaep,[2] Richard A. Flavell[1]*

Calcium is known to play vital roles in diverse physiological processes, and it is known that voltage-gated calcium channels (Ca_v) mediate calcium influx in excitable cells. However, no consensus exists on the molecular identity of the calcium channels present in nonexcitable cells such as T lymphocytes. Here, we demonstrate that T lymphocytes express both regulatory β_4 and pore-forming Ca_v1 α_1 subunits of Ca_v channels. Ca_v β_4-mutant T lymphocytes fail to acquire normal functions and display impairment in the calcium response, activation of the transcription factor NFAT, and cytokine production. Although Ca_v1 channels of lymphocytes retain their voltage dependency, T cell receptor stimulation dramatically increases channel opening, providing a new mechanism for calcium entry in lymphocytes.

Calcium plays critical and specific roles in many T cell functions, including activation, proliferation, and cytokine production (*1*–*3*). In T lymphocytes, ligation of the T cell receptor (TCR) by antigen leads to the release of calcium from intracellular stores, triggering the calcium release–activated calcium current (CRAC) (*4*). The molecular identity of CRAC channels is still unclear (*3*, *5*), although the complexity of the calcium response in T cells suggests the expression of more than one plasma membrane calcium channel. In other cell types, notably excitable cells, which take up calcium in response to membrane depolarization, Ca_v channels constitute the major route of calcium entry (*6*), although their potential role in nonexcitable cells, such as the T lymphocyte, is controversial (*7*).

Ca_v β subunits are cytoplasmic proteins that strongly regulate Ca_v channels through direct interaction with pore-forming α_1 subunits. The β subunits are also required for assembly of the channel complex (*8*), correct plasma membrane targeting (*9*, *10*), and stimulation of channel activity (*11*). A number of potential α_1-β combinations are likely to form a Ca_v channel complex (*12*); among these, the β_4 subunit is a major subunit associating with Ca_v1 channels (*13*, *14*).

A spontaneous mutation named *lethargic*, which arose in the mouse inbred strain BALB/cGn in 1962, is recognizable in homozygous mice at 2 weeks of age by onset of ataxia, seizures, and lethargic behavior (*15*, *16*). These mice also experience a generalized immunological disorder, including defective cell-mediated immune responses (*17*). Positional cloning revealed that the syndrome was the result of a mutation of the Ca_v β_4 subunit gene (*18*). A 4-nucleotide insertion into a splice donor site in lethargic mice has resulted in exon skipping and translational frameshift with loss of the α_1 subunit-binding site and a dramatic reduction of β_4 expression (*18*).

Semiquantitative reverse transcription polymerase chain reaction (RT-PCR) (Fig. 1A) and real-time quantitative PCR (Fig. 1B) revealed β_4 expression in unstimulated naïve T lymphocyte of wild-type mice, as well as up-regulation after TCR engagement. Because Ca_v β subunits can potentially interact with all Ca_v channels through an α_1-β interaction involving highly conserved residues in both subunits (*19*, *20*), we also determined which Ca_v channel subtypes are expressed by T lymphocytes before and after TCR stimulation (Fig. 1C). Three Ca_v channels belonging to the Ca_v1 family were detected: $Ca_v1.1$, $Ca_v1.2$, and $Ca_v1.4$, displaying differential expression as T lymphocytes transit from naive (day 0) toward an effector stage (day 4 after stimulation) (Fig. 1C). Ca_v2 and Ca_v3 channels were not detected either before or after stimulation (Fig. 1C). RT-PCR and Western blot assays also revealed a profound reduction of the expression of the β_4 subunit $Ca_v1.1$ but not $Ca_v1.2$ channels in Ca_v β_4-mutant CD4 T cells (Fig. 1D).

The functional activity of Ca_v β_4-mutant T cells was then tested using spleen and lymph node T cells from wild-type and β_4-mutant mice (fig. S1). The analysis of the thymus showed normal intrinsic T cell development. In fact, when mice were analyzed before the onset of the neurological syndrome (less than 2 weeks old), T cell development was normal (fig. S2A). In contrast, mice exhibiting the neuropathy (older than 2 weeks) showed a reduction in CD4/CD8 double-positive population (*21*). Together, the data suggest that this reduction is secondary to the neuropathy, including the corticosterone hypersecretion previously described in the mutant mice (*22*). This was confirmed using $RAG1^{-/-}$ bone marrow chimera mice to allow stem cells from both wild-type and β_4-mutant to develop in similar physiological environments. In these conditions, thymic development was indeed normal (fig. S2B). In addition, no difference in the percentage of CD4 and CD8 cells or in the expression of different activation markers (CD25, CD69, CD62L, and CD44) was observed in the mutant relative to wild-type littermate mice (fig. S2C), which suggests that the overall development of T cells is normal in these mice. Although typical calcium responses were obtained with wild-type cells (Fig. 2A, left panel), both the initial peak and the plateau were attenuated in the

[1]Section of Immunobiology, Howard Hughes Medical Institute, [2]Department of Cellular and Molecular Physiology, [3]Department of Pharmacology, Yale University School of Medicine, New Haven, CT 06510, USA. [4]Digestive Diseases Unit, Department of Medicine, Pharmacology and Physiology, University of Rochester School of Medicine.

*To whom correspondence should be addressed. E-mail: richard.flavell@yale.edu

β_4-mutant CD4 T lymphocytes (Fig. 2A, right panel, and Fig. 2B). Cells from both groups were able to respond similarly to the calcium ionophore ionomycin (Fig. 2A). The ability of both groups of cells to take up the calcium probe fluo-3 was similar between cells from wild-type and mutant mice (Fig. 2C), which rules out the possibility that the difference observed in the calcium response is secondary to a difference in the ability of the two groups of cells to take up fluo-3. Furthermore, no defect in the calcium response was observed when wild-type and mutant CD4+ T lymphocytes were stimulated with thapsigargin, which mediates a passive release of calcium from intracellular stores (*3*, *23*), suggesting that store-operated channels were normal in the β_4-mutant lymphocytes (Fig. 2D). No defect was observed in the release of calcium from intracellular stores as assessed by using TCR and thapsigargin stimulations in a calcium-free medium (fig. S3). These data show that the Ca_v β_4 regulatory subunit is necessary for a normal TCR-mediated calcium response in CD4 T cells.

An increase in $[Ca^{2+}]_i$ in lymphocytes leads to nuclear translocation of the transcription factor NFAT after its dephosphorylation by the calcium-dependent phosphatase calcineurin (*24*, *25*). This calcium/NFAT pathway is critical for T cell functions, including cytokine production. Nuclear translocation of both NFAT subtypes—NFATc2 and NFATc1—was inhibited in β_4-mutant lymphocytes (Fig. 2E), demonstrating that the complete calcium/NFAT pathway is substantially impaired in the absence of the β_4 regulatory subunit. Interestingly, a similar link between Ca_v1 channels and NFATc4 was reported in the hippocampal neurons (*26*).

Despite the strong effects of the β_4 subunit on calcium signaling, β_4-mutant CD4+ T lymphocytes did retain their ability to proliferate when stimulated via the TCR (Fig. 3A). TCR-mediated IL-2 production was partially inhibited in the mutant cells (Fig. 3B). The production of effector T cell cytokines, both IFNγ and IL-4, was also reduced in the β_4-mutant T cells under both primary and secondary conditions (Fig. 3C).

To exclude the possibility that the immune cell phenotype was secondary to the neuronal/endocrine deficiency, we generated bone-marrow chimeras using wild-type or β_4-mutant bone marrow transfer into irradiated RAG1-deficient mice. The analysis of CD4+ T cells purified from the chimeras showed a similar defect in IFNγ and IL-4 production (fig. S4), corroborating the findings observed with T cells from the β_4-mutant mice (Fig. 3). Thus, the T cell deficiency observed in β_4-mutant mice was intrinsic to T lymphocytes.

The TCR-mediated, β_4 subunit–dependent, calcium/NFATc1/NFATc2 pathway activation is therefore critical for cytokine production but is dispensable for proliferation of T lymphocytes. A selective defect in cytokine production but not T cell proliferation was also reported in NFATc2/NFATc1 double-knockout T cells (*27*). Our phenotype is also analogous to a report showing that two distinctive programs of gene expression are triggered to control the fiber type in the skeletal muscle, depending on the activation state of calcineurin (*28*, *29*).

In whole-cell patch-clamp studies, using Ba^{2+} instead of Ca^{2+} as a charge carrier, we evaluated whether Ca_v1 channels found in T lymphocytes are sensitive to depolarization. Voltage-dependent Ba^{2+} currents (I_{Ba}) were activated at approximately –35 mV (Fig. 4A, right panel), and maximum inward currents

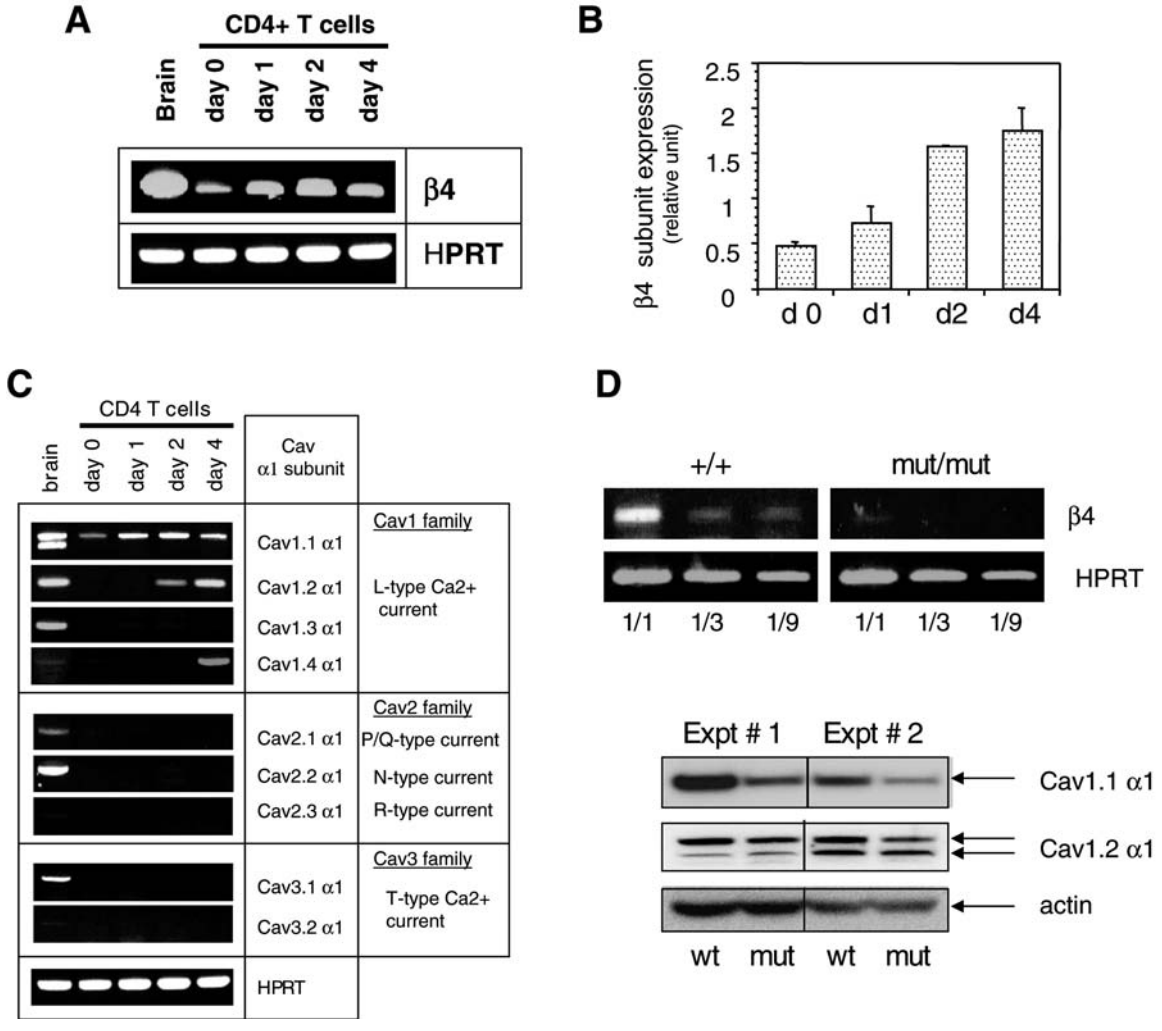


Fig. 1. Expression of Ca_v channel subunits by CD4+ T cells. Purified CD4 T cells were left unstimulated (day 0) or were stimulated with plate-bound antibodies to CD3 and CD28 for the indicated time. Brain cDNA was used as a positive control. (**A**) mRNA detection of the β_4 subunit by semiquantitative RT-PCR in CD4+ T cells. (**B**) β_4 subunit mRNA expression by CD4+ T cells confirmed with real-time quantitative PCR. (**C**) Expression of the Ca_v channel α_1 subunits by CD4+ T cells. (**D**) Expression of Ca_v channel subunits by β_4-mutant T cells assessed with RT-PCR and Western blot. Results are representative of at least three independent experiments, except for (D), in which two experiments were performed.

of 6.1 ± 0.53 pA/pF (n = 8 cells) were observed at +10 mV (Fig. 4A, right panel). This current inactivated slowly over time (Fig. 4A, left panel), consistent with the properties of Ca_v1 channels. The addition of the Ca_v1 channel blocker nifedipine (2 μM) decreased whole-cell I_{Ba} at all potentials positive to –35 mV (Fig. 4A, right panel). To further ensure that our measurements specifically re-

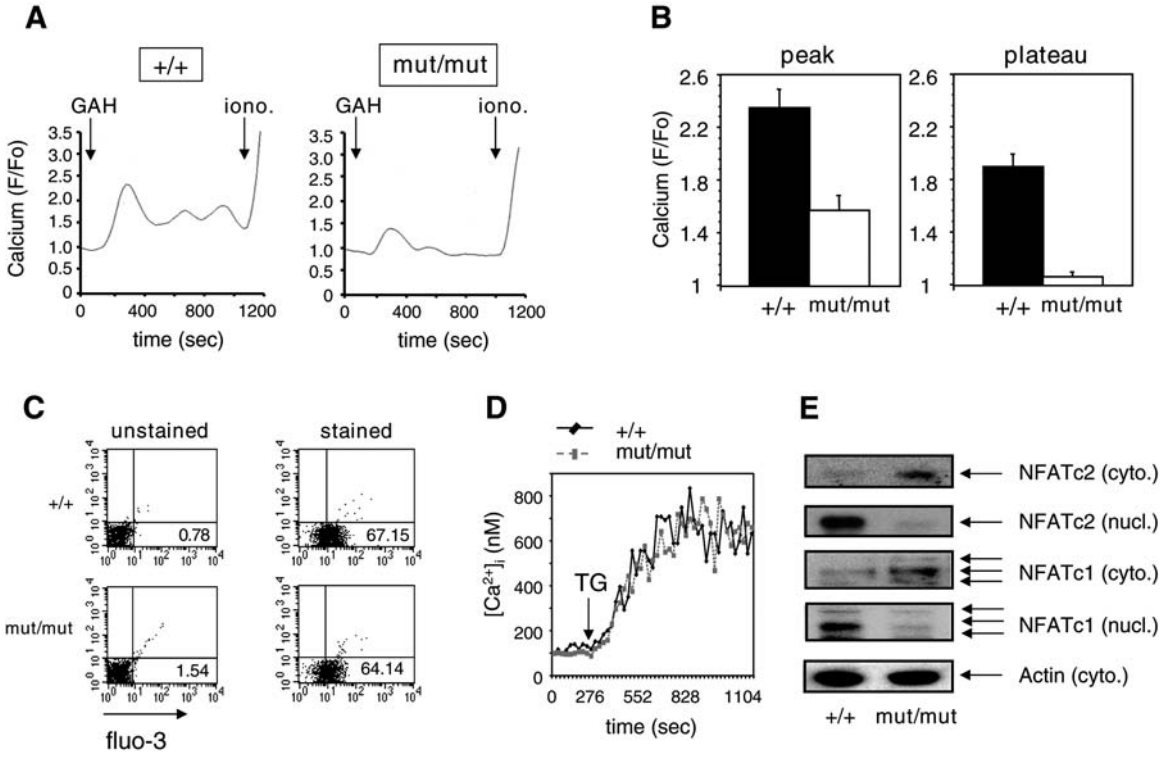


Fig. 2. The calcium/NFAT pathway is impaired in β_4-mutant CD4+ T cells. (**A**) The calcium response of CD4+ T cells from β_4-mutant mice (mut/mut) and wild-type (WT) littermate (+/+) was evaluated by using 2C11 antibody to CD3 and goat antihamster (GAH) in a cross-linking system. (**B**) Comparison of peak and plateau calcium responses of β_4-mutant T cells and control littermate. (**C**) Comparison of fluo-3/AM loading in T cells from β_4-mutant mice and WT littermate. (**D**) Calcium response of WT and β_4-mutant CD4+ T lymphocytes upon stimulation with Thapsigargin (1 μM). (**E**) Western blot assay was performed on cytoplasmic (cyto.) and nuclear (nucl.) extracts prepared from β_4-mutant CD4+ T cells and control littermate stimulated for 48 hours in the presence of human IL-2. Actin was used as an internal control. Results are representative of five independent experiments for (A) and (B), three for (D) and (E), and two for (C).

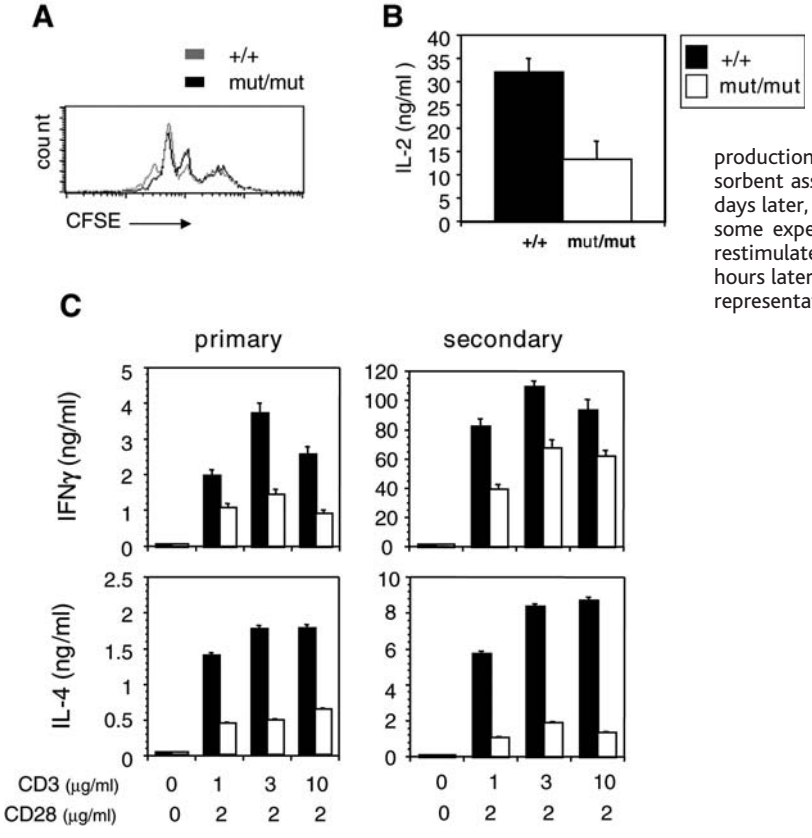


Fig. 3. Defective CD4+ T cell differentiation in the absence of functional β_4 subunit. (**A**) Proliferation of β_4-mutant CD4+ T cells stimulated by plate-bound antibodies to CD3 and CD28 assessed by carboxyfluorescein diacetate succinimidyl ester. (**B**) CD4+ T cells were stimulated as in (A) in the presence of exogenous human IL-2 (20 U/ml). IL-2 production was evaluated 24 hours later by enzyme-linked immunosorbent assay (ELISA). (**C**) CD4+ T cells were stimulated as in (A). Three days later, IFNγ and IL-4 production was measured by ELISA (primary). In some experiments, cells were stimulated for 4 days, then washed and restimulated (secondary) by plate-bound antibody to CD3. Twenty-four hours later, IFNγ and IL-4 production was measured by ELISA. Results are representative of at least three independent experiments.

flected a calcium channel, we replaced Ba^{2+} with 20 mM Ca^{2+}, which yielded a maximal current of 1.6 ± 0.23 pA/pF (Fig. 4B, n = 5 cells) and a degree of inhibition of 63 ± 1.8% in the presence of nifedipine (Fig. 4B).

Unitary Ca^{2+} currents were recorded in the cell-attached mode using pipettes containing 100 mM Ba^{2+}. Cells were held at their resting membrane potential and depolarized for 2.5 s. Single-channel currents were observed in only 35% of patched cells after depolarization (*30*). However, TCR stimulation dramatically increased the opening of the Ca_v1 channels, and currents were readily detected in 100% of the cells after depolarization (Fig. 4C). These findings reveal a direct link between the TCR and Ca_v1 channels in lymphocytes.

Further experiments were performed to characterize the electrical properties of single Ca_v1 channels. The Ca_v1 calcium channel agonist FPL64176 was included in the pipette using the same protocol as in Fig. 4C. The mean ensemble currents (I_{mean}) obtained by averaging 20 or more single traces revealed a time- and voltage-dependent inward current. Channel opening [NP(O), number of channels × open probability] increased at more depolarized potentials (Fig. 4E). The analysis of the single-channel current-voltage (*I-V*) relationship (Fig. 4F) indicated a unitary conductance of ~10.6 pS, similar to that previously reported for the $Ca_v1.1$ channel (*31*). No inward currents were detected in 10 of 10 patches on β_4-mutant T cells, even in the presence of FPL64176 (*32*).

Both whole-cell and single-channel recordings corroborate the presence of functional and TCR-gated Ca_v1 channels in T lymphocytes. These recordings show that the Ca_v1 channels expressed in T cells still retain their intrinsic ability to respond to voltage. To further assess the physiological role of depolarization in Ca_v1 channel opening in T lymphocytes, we monitored changes in the

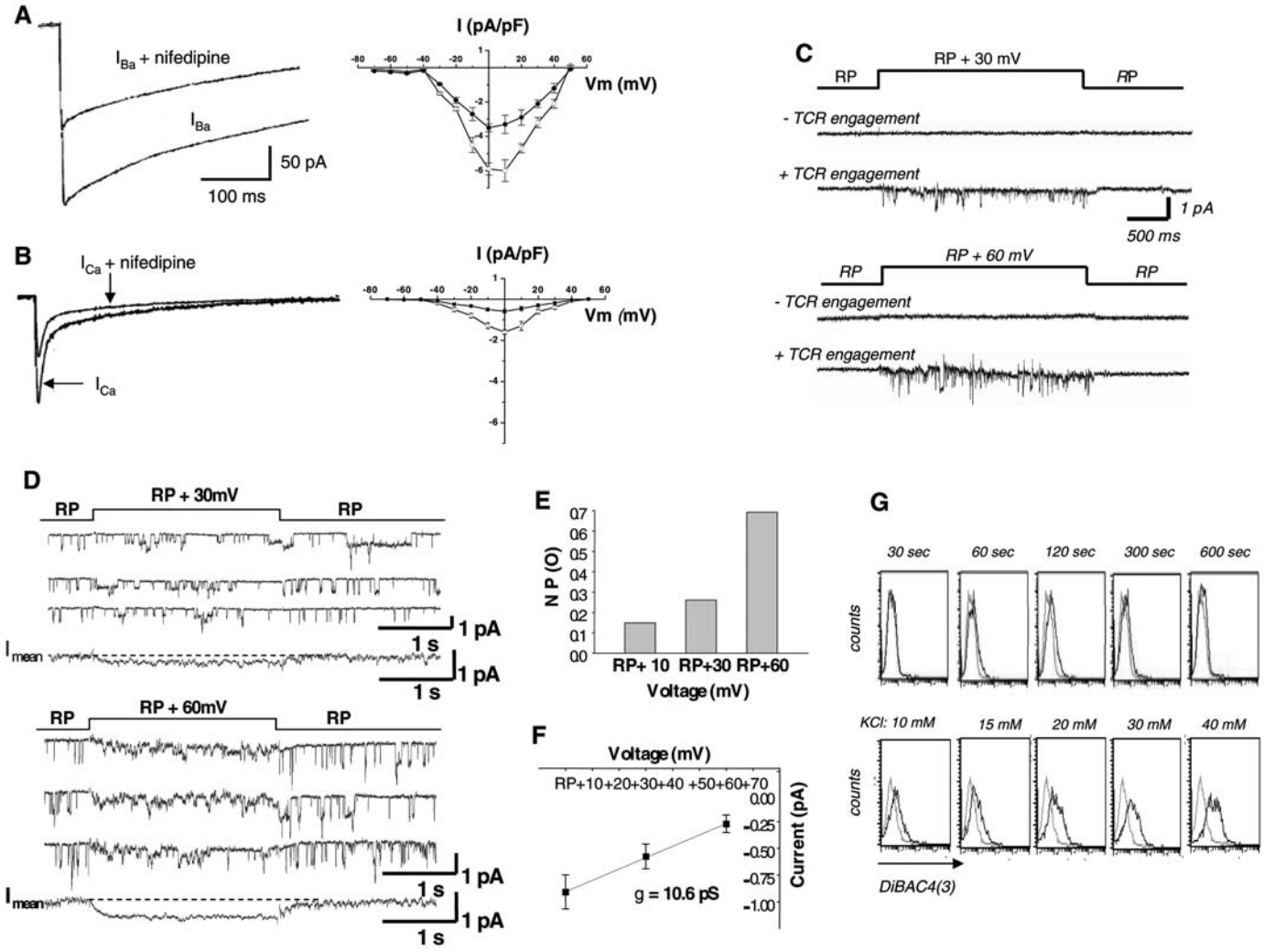


Fig. 4. Ca_v1 channels in T lymphocytes are voltage dependent. (**A**) The left panel shows a representative whole-cell current with Ba^{2+} as a charge carrier under basal conditions and after the addition of nifedipine (2 μM) in CD4+ effector T cells generated after stimulation with antibodies to CD3 and CD28 for 4 days. The right panel shows the current-voltage relation under basal conditions and after exposure to 2 μM of nifedipine. (**B**) Experiments similar to (A) were performed, but Ca^{2+} was used as a charge carrier. (**C**) Unitary currents were recorded in cell-attached patches in response to 2.5 s step depolarizations to potentials 30 mV and 60 mV positive to the resting membrane potential (RP) of the cell, using 100 mM Ba^{2+}. TCR cross linking was performed as indicated in Fig. 2. (**D**) Single-channel recordings were performed as in C but in the presence of the Ca_v1 channel agonist FPL64176 in the patch pipette. The occurrence of many channel openings at the resting potential may result, in part, from the action of the agonist. The mean ensemble current is shown below the traces (I_{mean}). Traces were corrected for leakage currents and capacitive transients were eliminated. (**E**) Bar graphs of the probability of opening for the channels shown in (D). (**F**) Single-channel *I-V* relationship for channels shown in (D). Linear regression yielded a unitary conductance of 10.6 pS. (**G**) The upper panel shows a modest depolarization of T lymphocytes after TCR stimulation (in black). Depolarization was assessed with the voltage-sensitive dye $DiBAC_4(3)$ by using flow cytometric analysis. The lower panel shows depolarization with different concentrations of KCl (10 to 40 mM). Results shown in (G) are representative of four independent experiments.

resting potential of T cells after TCR stimulation using the voltage-sensitive dye DiBAC4(3). A small depolarization was detected in stimulated T cells relative to unstimulated T cells (Fig. 4G, upper panels). This depolarization peaked at 120 s. A gradual repolarization was then observed 300 and 600 s after TCR stimulation (Fig. 4G, upper panels). The peak depolarization observed was similar to the depolarization obtained with 10 to 15 mM of KCl (Fig. 4G, lower panels) and was nifedipine sensitive (fig. S5), which suggests that Ca_v channels contribute to the generation/maintenance of this depolarization. This small depolarization might play a role in regulating Ca_v1 channels. Alternatively, Ca_v1 channels could be regulated in a voltage-independent manner, in which other TCR-triggered signaling events such as phosphorylation would induce opening. This hypothesis is strongly supported by the earlier unexpected findings that $Ca_v1.3$ channels can activate at voltages of $\sim$ –60mV under physiological calcium concentrations (*33*, *34*) and our observation of 100% single-channel opening after TCR ligation (Fig. 4C). Further studies are needed, however, to clarify the contribution of regulatory events other than voltage in Ca_v1 channel physiological gating in T lymphocytes.

The present study provides an answer to a long-standing question about the molecular nature of calcium channels present in T lymphocytes. We present strong genetic evidence illustrating the role of the β_4 regulatory subunit of Ca_v channels in T lymphocyte functions. The T cell functional defect described in the present work in the β_4-mutant T cells reflects the contribution of this subunit to Ca_v1 channel-dependent calcium influx in T lymphocytes. It is not yet clear which Ca_v1 is responsible for calcium entry in our studies. The fact that $Ca_v1.1$ levels were reduced (Fig. 1D) as a consequence of the β_4 mutation may contribute to the profound deficiency in these mice. It now remains to determine the specific role of each of these individual Ca_v1 channels.

References and Notes

1. D. Cantrell, *Annu. Rev. Immunol.* **14**, 259 (1996).
2. R. E. Dolmetsch, R. S. Lewis, C. C. Goodnow, J. I. Healy, *Nature* **386**, 855 (1997).
3. R. S. Lewis, *Annu. Rev. Immunol.* **19**, 497 (2001).
4. A. Zweifach, R. S. Lewis, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6295 (1993).
5. A. C. Elliott, *Cell Calcium* **30**, 73 (2001).
6. W. A. Catterall, *Annu. Rev. Cell Dev. Biol.* **16**, 521 (2000).
7. G. Grafton, L. Thwaite, *Immunology* **104**, 119 (2001).
8. E. Tareilus *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1703 (1997).
9. A. J. Chien, T. Gao, E. Perez-Reyes, M. M. Hosey, *J. Biol. Chem.* **273**, 23590 (1998).
10. T. Gao, A. J. Chien, M. M. Hosey, *J. Biol. Chem.* **274**, 2137 (1999).
11. D. Freise *et al.*, *Biol. Chem.* **380**, 897 (1999).
12. O. Tanaka, H. Sakagami, H. Kondo, *Brain Res. Mol. Brain Res.* **30**, 1 (1995).
13. M. Pichler *et al.*, *J. Biol. Chem.* **272**, 13877 (1997).
14. J. M. Schjott, S. C. Hsu, M. R. Plummer, *J. Biol. Chem.* **278**, 33936 (2003).
15. M. M. Dickie, *Mouse News Lett.* **30**, 31 (1964).
16. R. L. Sidman, M. C. Green, S. H. Appel, *Catalog of the Neurological Mutants of the Mouse* (Cambridge, MA, Harvard Univ. Press, 1965).
17. D. G. Morrison, M. P. Moyer, H. C. Dung, W. Rogers, R. C. Moyer, *Dev. Comp. Immunol.* **8**, 435 (1984).
18. D. L. Burgess, J. M. Jones, M. H. Meisler, J. L. Noebels, *Cell* **88**, 385 (1997).
19. M. Pragnell *et al.*, *Nature* **368**, 67 (1994).
20. M. De Waard, M. Pragnell, K. P. Campbell, *Neuron* **13**, 495 (1994).
21. A. Badou *et al.*, unpublished observation.
22. H. C. Dung, *Am. J. Anat.* **147**, 255 (1976).
23. E. Mintz, F. Guillain, *Biochim. Biophys. Acta* **1318**, 52 (1997).
24. A. Rao, C. Luo, P. G. Hogan, *Annu. Rev. Immunol.* **15**, 707 (1997).
25. G. R. Crabtree, E. N. Olson, *Cell* **109** (suppl), S67 (2002).
26. I. A. Graef *et al.*, *Nature* **401**, 703 (1999).
27. S. L. Peng, A. J. Gerth, A. M. Ranger, L. H. Glimcher, *Immunity* **14**, 13 (2001).
28. E. R. Chin *et al.*, *Genes Dev.* **12**, 2499 (1998).
29. Y. Liu, Z. Cseresnyes, W. R. Randall, M. F. Schneider, *J. Cell Biol.* **155**, 27 (2001).
30. A. Badou *et al.*, unpublished observation.
31. D. Freise *et al.*, *J. Biol. Chem.* **275**, 14476 (2000).
32. A. Badou *et al.*, unpublished observation.
33. A. Koschak *et al.*, *J. Biol. Chem.* **276**, 22100 (2001).
34. W. Xu, D. Lipscombe, *J. Neurosci.* **21**, 5944 (2001).

35. The authors thank F. Manzo for assistance with manuscript preparation. A.B. was supported by the Fondation pour la Recherche Médicale and is now supported by the Arthritis National Research Foundation. S.B. is supported by an American Gastrointestinal Association (AGA) Research Scholar Award. D.M. is supported by a Cancer Research Institute postdoctoral fellowship. A.B. was an Associate and R.A.F. is an Investigator of the Howard Hughes Medical Institute.

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5706/117/DC1
Materials and Methods
Figs. S1 to S5
References

21 May 2004; accepted 23 November 2004
10.1126/science.1100582

The Enigma of Prokaryotic Life in Deep Hypersaline Anoxic Basins

Paul W. J. J. van der Wielen,[1]*† Henk Bolhuis,[1] Sara Borin,[2] Daniele Daffonchio,[2] Cesare Corselli,[3] Laura Giuliano,[4] Giuseppe D'Auria,[4] Gert J. de Lange,[5] Andreas Huebner,[5] Sotirios P. Varnavas,[6] John Thomson,[7] Christian Tamburini,[8] Danielle Marty,[8] Terry J. McGenity,[9] Kenneth N. Timmis,[9,10] BioDeep Scientific Party

Deep hypersaline anoxic basins in the Mediterranean Sea are a legacy of dissolution of ancient subterranean salt deposits from the Miocene period. Our study revealed that these hypersaline basins are not biogeochemical dead ends, but support in situ sulfate reduction, methanogenesis, and heterotrophic activity. A wide diversity of prokaryotes was observed, including a new, abundant, deeply branching order within the *Euryarchaeota*. Furthermore, we demonstrated the presence of a unique, metabolically active microbial community in the Discovery basin, which is one of the most extreme terrestrial saline environments known, as it is almost saturated with $MgCl_2$ (5 M).

Deep hypersaline anoxic basins (hypersaline basins) found in the Eastern Mediterranean Sea (*1*–*5*) probably resulted from the dissolution of subterranean Miocene salt deposits that became exposed to seawater after tectonic activity (*1*). Brines enclosed in these basins are characterized by anoxic conditions, high pressure (around 35 MPa), and almost saturated salt concentrations (*1*–*5*). The high densities of the hypersaline brines limit mixing with overlaying oxic seawater and result in a sharp chemocline of 1 to 3 m. It is clear from this and previous studies that each hypersaline basin is geochemically distinct (table S1) (*1*, *3*–*10*). L'Atalante, Bannock, and Urania brines have similar dominant ion compositions, but the overall salinity of Urania is lower, whereas concentrations of sulfide and methane are considerably higher (table S1). The most striking difference between the geochemistry of Discovery brine compared with the other three is the extremely high concentration of Mg^{2+} and low concentration of Na^+ (table S1). The physical separation of the basins from each other, as well as their existence for thousands of years, may have resulted in the evolution of specific microbial communities in each hypersaline basin. To date, only a few Bacteria, typically found in oxic seawater, have been isolated from the chemocline (*11*–*14*), but it remains unknown whether they are active in the hypersaline basins. The Discovery Basin contains a brine that has the highest concentration of $MgCl_2$ ($\sim$5 M) found thus far in a marine environment (*5*, *8*); such concentrations are considered anathema to life.

We embarked on a detailed study of four different hypersaline basins: L'Atalante, Bannock, Discovery, and Urania, to investigate their geochemistry, uncultivated microbiota,

and in situ microbial metabolic activities. Our aims were to determine the extent to which geochemical conditions influence the evolution of brine communities and to study whether life is possible under the hostile conditions of Discovery brine.

Microbial cells stained with 4′,6-diamidino-2-phenylindole (DAPI) were observed in the four brines with numbers ranging from 1.9×10^4 ml^{-1} in Discovery to 1.5×10^5 ml^{-1} in Urania (table S1) (*15*). The Bacteria/Archaea ratio, based on fluorescence in situ hybridization, indicated that Bacteria dominated the Discovery basin and were slightly more abundant in L'Atalante and Bannock basin, whereas Archaea dominated the Urania basin. In all four hypersaline basins, bacterial diversity was higher than archaeal diversity (table S1), with Urania basin showing lowest overall diversity. All of the hypersaline basins, including Discovery basin, showed evidence of sulfate reduction, methanogenesis, and heterotrophic activity (table S1). Analysis of the 16*S* ribosomal RNA (rRNA) gene sequence showed that high percentages of clone sequences obtained from the basins belonged to γ-, δ-, and ε-*Proteobacteria*; *Sphingobacteria*; candidate division KB1 (*16*); *Halobacteria*; and a new division which we named MSBL1 [Mediterranean Sea Brine Lakes group 1 (table S1)]. In contrast, seawater above these hypersaline basins showed a very different community structure (table S1). Typical activities of representatives of most of the phylogenetic groups found in the hypersaline basins suggest that they are responsible for the observed sulfate reduction and heterotrophic activity. Thus, deep hypersaline anoxic basins of the Mediterranean are not biological dead-ends but contain active microbial communities that contribute to biogeochemical cycling of carbon and sulfur, as has been observed in other anoxic, highly saline ecosystems (*17*). Methane was produced in all four hypersaline basins (table S1), but few 16*S* rRNA gene sequences related to known methanogenic Archaea were found. The majority of the archaeal 16*S* rRNA gene sequences discovered belonged to a new subdivision that branched deeply within the *Euryarchaeota*, candidate division MSBL1 (Fig. 1). This candidate division is equivalent in genetic depth and breadth to, for example, *Halobacteriales* and represents a new order of yet-to-be-cultivated Archaea. On the basis of phylogenetic relatedness of the MSBL1 Archaea to methanogens and the lack of any other group detected that might be responsible for the methane production, it is reasonable to speculate that MSBL1-related Archaea are involved in methanogenesis at high salinity.

The observations that Discovery brine contains 1.9×10^4 microbial cells ml^{-1} and that most of the 16*S* rRNA gene sequences are related to phylogenetic groups not found in normal seawater indicate that a specific microbial community is present. Other environments with elevated concentrations of $MgCl_2$ are the Dead Sea and Lake Bonney in the Antarctic (*18*, *19*), but concentrations do not exceed 2 M. Haloarchaeal species capable of growth at 1 M $MgCl_2$ have been isolated from the Dead Sea (*20*, *21*). However, life has never been observed in, nor have prokaryotes been isolated, that can tolerate 5 M $MgCl_2$, the concentration in Discovery hypersaline brine (table S1).

The bacterial community of the Discovery brine was unequivocally different from that of the Discovery interface and overlying seawater, as indicated by automated ribosomal intergenic spacer analysis profiles (*22*). Similar results were obtained when Morisita

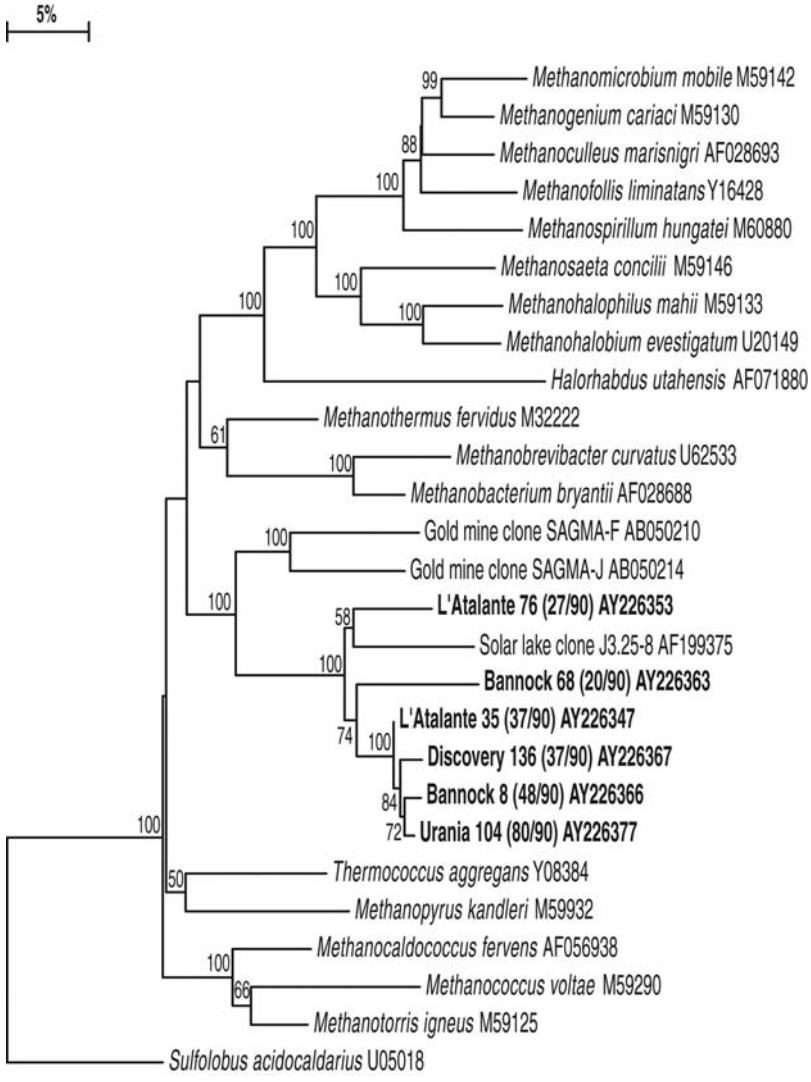


Fig. 1. Phylogenetic tree showing the most dominant archaeal sequences belonging to the new subdivision MSBL1 from the four different deep hypersaline anoxic basins. Trees were constructed with the neighbor-joining method using 100 bootstrap replicates. The values between parentheses show the number of the specific sequence found compared with the total number of archaeal sequences analyzed. Bootstrap values above 50% are shown. Scale bar, 5% of evolutionary distance.

[1]Laboratory of Microbial Ecology, University of Groningen, 9751 NN Haren, Netherlands. [2]Consorzio Nazionale Interuniversitario per le Scienze del Mare (CoNISMa), Università degli Studi di Milano, URL: DiSTAM, 20133 Milan, Italy. [3]CoNISMa, Università degli Studi di Milano Bicocca, URL: DGSG, 20126 Milan, Italy. [4]Istituto per L'Ambiente Marino Costiero, CNR, 98122 Messina, Italy. [5]Faculty of Geosciences, Geochemistry, Utrecht University, 3508 TA Utrecht, Netherlands. [6]Department of Geology, University of Patras, 26500 Patras, Greece. [7]Southampton Oceanography Centre, Southampton, SO14 3ZH, UK. [8]Laboratoire de Microbiologie, Géochimie & Ecologie des Marines (LMGEM), UMR 6117 CNRS, Université de la Méditerranée, 13 288 Marseille Cedex 9, France. [9]Department of Biological Sciences, University of Essex, Colchester, Essex CO4 3SQ, UK. [10]Division of Microbiology, Gesellschaft für Biotechnologische Forschung mbH (GBF), D-38124 Braunschweig, Germany.

*Present address, Kiwa Water Research, Post Office Box 1072, 3430 BB Nieuwegein, Netherlands.
†To whom correspondence should be addressed. E-mail: paul.van.der.wielen@kiwa.nl

indices of similarity (*23*) were calculated for the archaeal and bacterial operational taxonomic unit (OTU) distribution, based on 16*S* rRNA gene sequence analysis. The similarity indices of bacterial or archaeal OTU distribution between seawater and the seawater-brine interface were 0.001 and 0.01, respectively; these values were 0.48 for Bacteria and 0.89 for Archaea when OTU distributions between interface and brine were compared. Clearly, the seawater communities were different from the communities occurring in the Discovery interface or brine, whereas the brine communities did include microorganisms found at the interface. However, most bacterial sequences and part of the archaeal sequences were unique to the hypersaline Discovery brine. Most striking is the proportional increase, from seawater to brine, of clone sequences similar to *Halorhabdus utahensis* (99% similarity), an Archaeon that tolerates up to 0.8 M $MgCl_2$ (*24*). It was absent in the archaeal clone library of seawater, but constituted 11% of the interface clone libraries and, strikingly, 33% of the brine clone libraries (table S1). Ectoenzymatic activities, glutamic acid uptake, sulfate reduction, and methane production rates were measured at in situ temperature in brine and interface samples (table S1). These results show that prokaryotes can be active under the extreme conditions of Discovery basin. The rates of ectoenzymatic activities were higher in Discovery compared with the three NaCl-dominated basins. In contrast, methane production and sulfate reduction rates were lower in Discovery than in the other basins, except for the sulfate reduction rate in Bannock. This might indicate that in Discovery brine heterotrophic prokaryotes are more important than those involved in methanogenesis and sulfate reduction compared with the other brines. This idea is supported by the observation that clone sequences of a phylotype like *H. utahensis*, a heterotrophic Archaeon capable of fermentative growth, is enriched relative to the MSBL1 group in Discovery when compared with L'Atalante, Bannock, and Urania brines. Results from our study show that life is possible in brines with very high concentrations of divalent cations and shed new light on the existence of microbial life in other types of brines. For example, it has been questioned whether life is possible in Don Juan Pond, an Antarctic brine of 3.2 M $CaCl_2$ (*19*, *25*, *26*).

A cluster analysis of the combined archaeal and bacterial 16*S* rRNA gene data from the four hypersaline basins highlighted the differences between the microbial communities present (Fig. 2, left). The microbial community in Discovery basin is most different from the others, whereas Bannock and L'Atalante are most similar. Cluster analysis of the four basins based on (bio)-geochemical data from table S1 showed a pattern identical to that of the 16*S* rRNA gene data, indicating that microbial community structure and geochemical conditions are directly linked (Fig. 2, right). The high concentration of $MgCl_2$ in the Discovery basin is likely to influence evolution of the microbial community in its brine, because most halophilic microorganisms studied can grow under high NaCl concentrations but not under high $MgCl_2$ concentrations (*27*). The elevated concentrations of sulfide and methane in the Urania basin provide valuable sources of energy for many microorganisms, so the higher concentrations of these substrates might have selected a microbial community with high numbers of sulfide- and methane-oxidizing microbes, as well as sulfide-tolerant prokaryotes.

Evidence that microbial life is possible at 5 M $MgCl_2$ widens the picture of microbial adaptation to salinity. It has been suggested that primordial life on earth started in hypersaline water (*28*, *29*); furthermore, extraterrestrial objects are known to contain brines exposed to evaporation, which results in an increase of divalent cations (*30*, *31*). Our results indicate that microbial metabolism can proceed at significant levels in some of the most extreme terrestrial hypersaline environments and lend further support to the possibility of extraterrestrial life.

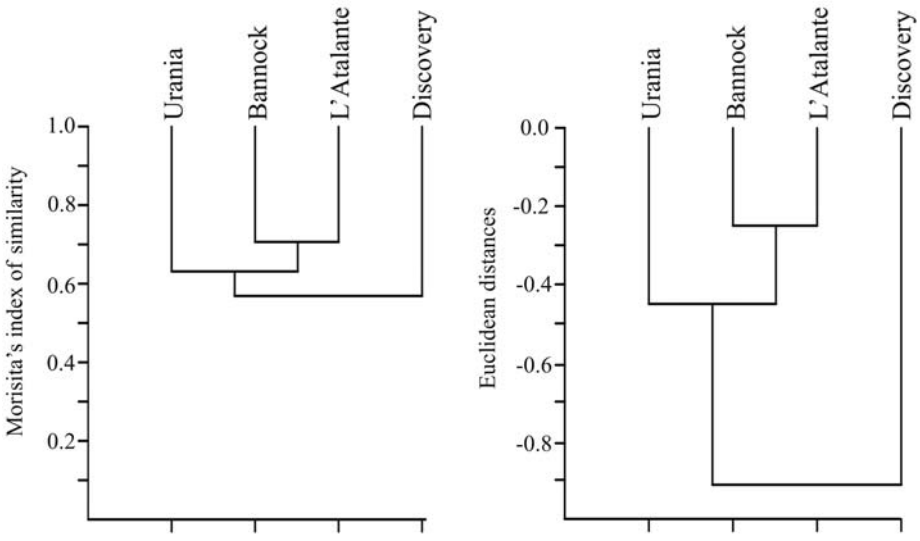


Fig. 2. UPGMA (unweighted pair group method with arithmetic mean) cluster analysis based on the Morisita index of similarity between the operational taxonomic unit (>97% 16*S* rRNA gene sequence similarity) distribution of the four deep hypersaline anoxic basins (**left**), and Euclidean distances between the geochemical data of table S1 for the four hypersaline basins (**right**).

References and Notes

1. A. Camerlenghi, *Mar. Chem.* **31**, 1 (1990).
2. D. Jongsma *et al.*, *Nature* **305**, 795 (1983).
3. G. J. de Lange *et al.*, *Mar. Chem.* **31**, 63 (1990).
4. MEDRIFF Consortium, *Eos* (*Washington, DC*) **76**, 313 (1995).
5. K. Wallmann, E. Suess, G. H. Westbrook, G. Winckler, M. B. Cita, *Nature* **387**, 31 (1997).
6. S. M. Karisiddaiah, *Deep Sea Res. I Oceanogr. Res. Pap.* **47**, 1999 (2000).
7. A. Vengosh, G. J. de Lange, A. Starinsky, *Geochim. Cosmochim. Acta* **62**, 3221 (1998).
8. K. Wallmann *et al.*, *Mar. Geol.* **186**, 9 (2002).
9. W. Ziebis *et al.*, abstract of a paper presented at Goldschmidt 2000, Oxford, UK, 3 to 8 September 2000, *J. Conf. Abstr.* **5** (2), 1134 (2000); available at www.the-conference.com/JConfAbs.html.
10. E. Henneke, G. J. de Lange, *Mar. Chem.* **31**, 113 (1990).
11. T. Brusa *et al.*, *Microbiol. Res.* **152**, 45 (1997).
12. T. Brusa *et al.*, *Microbiol. Res.* **156**, 49 (2001).
13. R. La Ferla, E. Crisafi, *Mar. Ecol. Prog. Ser.* **75**, 309 (1991).
14. A. M. Sass, H. Sass, M. J. L. Coolen, H. Cypionka, J. Overmann, *Appl. Environ. Microbiol.* **67**, 5392 (2001).
15. Materials and methods are available as supporting material on *Science* Online.
16. W. Eder, W. Ludwig, R. Huber, *Arch. Microbiol.* **172**, 213 (1999).
17. B. Ollivier, P. Caumette, J. L. Garcia, R. A. Mah, *Microbiol. Rev.* **58**, 27 (1994).
18. O. Matsubaya, H. Sakai, T. Torii, H. Burton, K. Kerry, *Geochim. Cosmochim. Acta* **43**, 7 (1979).
19. A. Oren, *Hydrobiology* **405**, 1 (1999).
20. A. Oren, *Int. J. Syst. Bacteriol.* **33**, 381 (1983).
21. A. Oren, P. Gurevich, R. T. Gemmell, A. Teske, *Int. J. Syst. Bacteriol.* **45**, 747 (1995).
22. P. W. J. J. van der Wielen *et al.*, unpublished observations.
23. H. Wolda, *Oecologia* **50**, 296 (1981).
24. M. Waino, B. J. Tindall, K. Ingvorsen, *Int. J. Syst. Evol. Microbiol.* **50**, 183 (2000).
25. N. E. Horowitz, R. E. Cameron, J. S. Hubbard, *Science* **176**, 242 (1972).
26. B. Z. Siegel, G. McMurty, S. M. Siegel, J. Chen, P. LaRock, *Nature* **280**, 828 (1979).
27. A. Oren, *FEMS Microbiol. Rev.* **13**, 415 (1994).
28. I. Dundas, *Extremophiles* **2**, 375 (1998).
29. L. P. Knauth, *Nature* **395**, 554 (1998).
30. E. J. Gaidos, F. Nimmo, *Nature* **405**, 637 (2000).
31. J. S. Kargel, *Icarus* **94**, 368 (1991).

32. This work was supported by the European Commission's Sustainable Marine Ecosystem program, under the BIODEEP project (contract EVK3-2000-22057). K.N.T. gratefully acknowledges the generous support of the Fonds der Chemischen Industrie.

Supporting Online Material
www.sciencemag.org/cgi/content/full/307/5706/121/DC1
Materials and Methods
Table S1
References and Notes

2 August 2004; accepted 5 November 2004
10.1126/science.1103569

Vesicle Endocytosis Requires Dynamin-Dependent GTP Hydrolysis at a Fast CNS Synapse

Takayuki Yamashita, Toshihide Hige, Tomoyuki Takahashi*

Molecular dependence of vesicular endocytosis was investigated with capacitance measurements at the calyx of Held terminal in brainstem slices. Intraterminal loading of botulinum toxin E revealed that the rapid capacitance transient implicated as "kiss-and-run" was unrelated to transmitter release. The release-related capacitance change decayed with an endocytotic time constant of 10 to 25 seconds, depending on the magnitude of exocytosis. Presynaptic loading of the nonhydrolyzable guanosine 5′-triphosphate (GTP) analog GTPγS or dynamin-1 proline-rich domain peptide abolished endocytosis. These compounds had no immediate effect on exocytosis, but caused a use-dependent rundown of exocytosis. Thus, the guanosine triphosphatase dynamin-1 is indispensable for vesicle endocytosis at this fast central nervous system (CNS) synapse.

Endocytosis of distinct kinetics and mechanisms are thought to operate in parallel at nerve terminals (*1*) and to be critically involved in the maintenance of synaptic transmission (*2*). However, there is no direct information about the molecular dependence of endocytosis at the nerve terminal. The monomeric guanosine triphosphatase (GTPase) dynamin is indispensable for *Drosophila* neuromuscular transmission (*3*, *4*) and is thought to play an essential role in the fission reaction of clathrin-coated vesicles during endocytosis (*5*). However, at goldfish retinal ribbon synapses, fast endocytosis depends entirely on adenosine triphosphatases (ATPases), not on GTPases (*6*). Thus, it is not known to what extent dynamin is essential for endocytosis at the CNS synapse. At hippocampal synapses (*7*, *8*) and secretory cells (*9*), the flickering fusion pore called "kiss-and-run" (*10*) has been proposed to serve as rapid vesicle retrieval. At the giant nerve terminal calyx of Held, a rapid capacitance change, observed with spontaneous or evoked synaptic responses (*11*), has been implicated as kiss-and-run (*12*). At this presynaptic terminal, a blockade of GTP hydrolysis causes a use-dependent rundown of excitatory postsynaptic currents (EPSCs), indicating that GTPases are essential for vesicle recycling (*13*). However, neither the type of GTPases involved nor the GTP-dependent step in the recycling pathway is clarified.

To address these questions, we made simultaneous pre- and postsynaptic recordings combined with capacitance measurements at the calyx of Held synapse in rat brainstem slices (*14*). Presynaptic Ca^{2+} currents, evoked by a depolarizing pulse, elicited EPSCs, increased presynaptic membrane capacitance, and transiently decreased presynaptic membrane resistance without changing series resistance (Fig. 1A) (*15*). To determine whether the capacitance change is correlated with synaptic transmission, we loaded botulinum toxin E (BoNT/E; 200 nM), which specifically synaptosome-associated protein of 25 kD (SNAP-25) (*16*), directly into the calyceal terminal through a patch pipette. Evoked EPSCs were gradually diminished and eventually abolished 10 to 15 min after loading BoNT/E at 32° to 34°C (Fig. 1B), whereas spontaneous miniature EPSCs remained. Concomitantly, the capacitance change was blocked but for a rapidly decaying component. This component (±SEM) was 64.3 ± 18.5 pF (n = 5) in amplitude and 227 ± 62 ms (n = 5) in weighted mean decay time constant, similar in both magnitude and kinetics to that implicated as kiss-and-run exo-endocytosis (*11*, *12*). This capacitance component remained stable in both kinetics and amplitude during repeated stimulation in the presence of BoNT/E (fig. S1A) (*17*), but was blocked by Cd^{2+} (0.25 mM, Fig. 1B) as previously reported (*11*). A similar BoNT/E-resistant capacitance change could also be evoked by a brief (1-ms) depolarizing pulse (*11*) (fig. S1B). These results suggest that the rapid capacitance change implicated as kiss-and-run is unrelated to synaptic transmission, but might arise from SNAP-25–independent nontransmitter or extrasynaptic vesicle exocytosis or from a nonvesicular artifact.

Given that the BoNT/E-resistant capacitance and resistance changes lasted 300 to 400 ms after depolarization, we restricted our analysis to the capacitance change 450 to 500 ms after depolarization (ΔC_{mt}) (mt indicates transmitter release–related membrane capacitance change). As the duration of depolarizing pulse was shortened, the magnitude of ΔC_{mt} decreased exponentially with the charge of Ca^{2+} current (Q_{Ca}) (Fig. 1C). ΔC_{mt} could also be evoked by a weak stimulation, which is comparable in Q_{Ca} to that evoked by an action potential waveform (*18*) (AP-eq, Fig. 1C). The ΔC_{mt} decayed with a single or double exponential time course; its time constant (τ_{endo}) became shorter in linear proportion to ΔC_{mt} (Fig. 1D). The τ_{endo} (±SEM) for ΔC_{mt} evoked by a 10-ms pulse was 24.7 ± 3.3 s (n = 8), compared with 10.4 ± 1.4 s (n = 6) for that evoked by AP-eq stimulation. These results are consistent with the previous reports that the time required for endocytosis is proportional to the amount of exocytosis (*11*, *19*, *20*).

We next tested whether guanine nucleotide analogs might affect exo- or endocytosis. Intraterminal loading of 3 mM guanosine 5′-*O*-(2-thiodiphosphate) (GDPβS), which inhibits GTPase activity (*13*), slowed τ_{endo} (10 ms) twofold (51.7 ± 7.6 s, n = 6, $P < 0.05$, Fig. 2A), whereas it had no effect on ΔC_{mt} evoked by the first stimulus (fig. S2A). During repeated 10-ms stimulation at 1- or 2-min intervals, the ΔC_{mt} amplitude decreased, more rapidly for the shorter interstimulus interval (Fig. 2B). These results are consistent with the finding that intraterminal loading of GDPβS has no effect on basal transmission, but slows recovery from short-term depression in a frequency-dependent manner (*13*), suggesting together that GTPase activity is essential for accelerating recruitment of synaptic vesicles.

Intraterminal loading of a 0.2 mM solution of the nonhydrolyzable GTP analog guanosine 5′-*O*-(3-thiotriphosphate) (GTPγS) completely abolished endocytosis after 10-ms stimulation, except for the early component (up to 4 s; Fig. 2A), and completely abolished that following AP-eq stimulation. When stimulated repeatedly by a 10-ms pulse, the ΔC_{mt} amplitude decreased in a use-dependent manner; the rate of decline was the same for the 1- or 2-min interstimulus interval (Fig. 2B). GTPγS had no clear effect on the initial amplitude of ΔC_{mt} (*21*) (fig. S2). These results indicate that GTP hydrolysis is indispensable for synaptic vesicle endocytosis at the calyx of Held.

We next examined whether dynamin might be involved in endocytosis at the calyx of Held. The proline-rich domain (PRD) peptide of dynamin-1 disrupts interactions between dynamin and amphiphysin (*22*). Injection of this peptide into a reticulospinal presynaptic axon of lamprey increases the number of unfissioned coated vesicles (*23*). When we

Department of Neurophysiology, University of Tokyo Graduate School of Medicine, Tokyo 113-0033, Japan.

*To whom correspondence should be addressed. E-mail: ttakahas-tky@umin.ac.jp

loaded the PRD peptide (1 mM) into the calyceal terminal, endocytosis following exocytosis induced by AP-eq stimulation was completely abolished and endocytosis following a 10-ms pulse was blocked, except for the early component (Fig. 3A), whereas ΔC_{mt} at the first stimulus was unchanged (fig. S2). Upon repetitive 10-ms stimulation, ΔC_{mt} underwent a use-dependent rundown (Fig. 3B). A scramble control peptide had no such effect (Fig. 3, A and B). Thus, the effect of PRD peptide was essentially the same as GTPγS, but for the relatively slower rundown of ΔC_{mt} compared with that of GTPγS. This difference may arise from other GTPases such as Rab3 (*24*) or Arf (*25*), which may also be involved in vesicle recycling steps.

To examine whether GTPase-independent recycling mechanisms might additionally operate at the calyx of Held, we carried out continuous 10-ms stimulation at 1-min intervals in the presence of GTPγS while monitoring ΔC_{mt} (Fig. 4A) or EPSCs (Fig. 4B). After 30 to 35 stimuli, ΔC_{mt} became undetectable (<10 fF, n = 4), whereas EPSCs were still observed (*26*). After 40 stimuli, however, EPSCs were completely abolished (n = 4). Because GTPγS reduces Ca^{2+} currents by activating trimeric G proteins (*27*), we compensated for it by increasing extracellular Ca^{2+} concentration (to 6 mM, with Mg^{2+} reduced to 0.1 mM). Even in this condition, no EPSC was evoked (Fig. 4B, n = 3), suggesting that vesicle recycling entirely depends upon GTP hydrolysis at this synapse (*28*).

From the total vesicles released during repetitive stimulation in the presence of GTPγS, the number of recycling vesicles (±SEM) can be estimated as 42,000 ± 6900 (n = 4) (*14*). This number is twice as large as that estimated from styryl dye–uptake experiments (*29*) and corresponds to 20% of total synaptic vesicles at the calyx of Held (*29*). Also, the maximal releasable pool size can be estimated, from the number of vesicles released by a 10-ms depolarizing pulse, as 4470 ± 400 (n = 14). This number, corresponding to 11% of recycling vesicles, is similar to those previously reported (*15*, *30*, *31*).

Synaptic vesicles fused into plasma membrane are reused through endocytosis and recycling (*1*, *2*). Among various recycling routes with distinct kinetics (*1*), a subsecond capacitance change evoked by a brief presynaptic Ca^{2+} current has been implicated as kiss-and-run exo-endocytosis at the calyx of Held (*11*, *12*). Although our results obtained with BoNT/E do not support this implication, they do not necessarily preclude the presence of dynamin-dependent kiss-and-run (*32*) at this nerve terminal. Furthermore, abolishment of endocytosis by GTPγS or dynamin-1 PRD peptide was incomplete when exocytosis was evoked by a 10-ms depolarizing pulse, whereas it was complete when evoked by AP-eq stimulation. Although our results suggest that the dynamin-1–dependent vesicular fission mechanism predominantly mediates endocytosis at the calyx of Held, there may additionally be a dynamin-independent endocytotic mechanism such as bulk endocytosis (*1*), which may operate after massive exocytosis (*29*). However, complete block of transmitter release after prolonged stimulation in the presence of GTPγS indicates that GTP hydrolysis is indispensable for vesicle recycling.

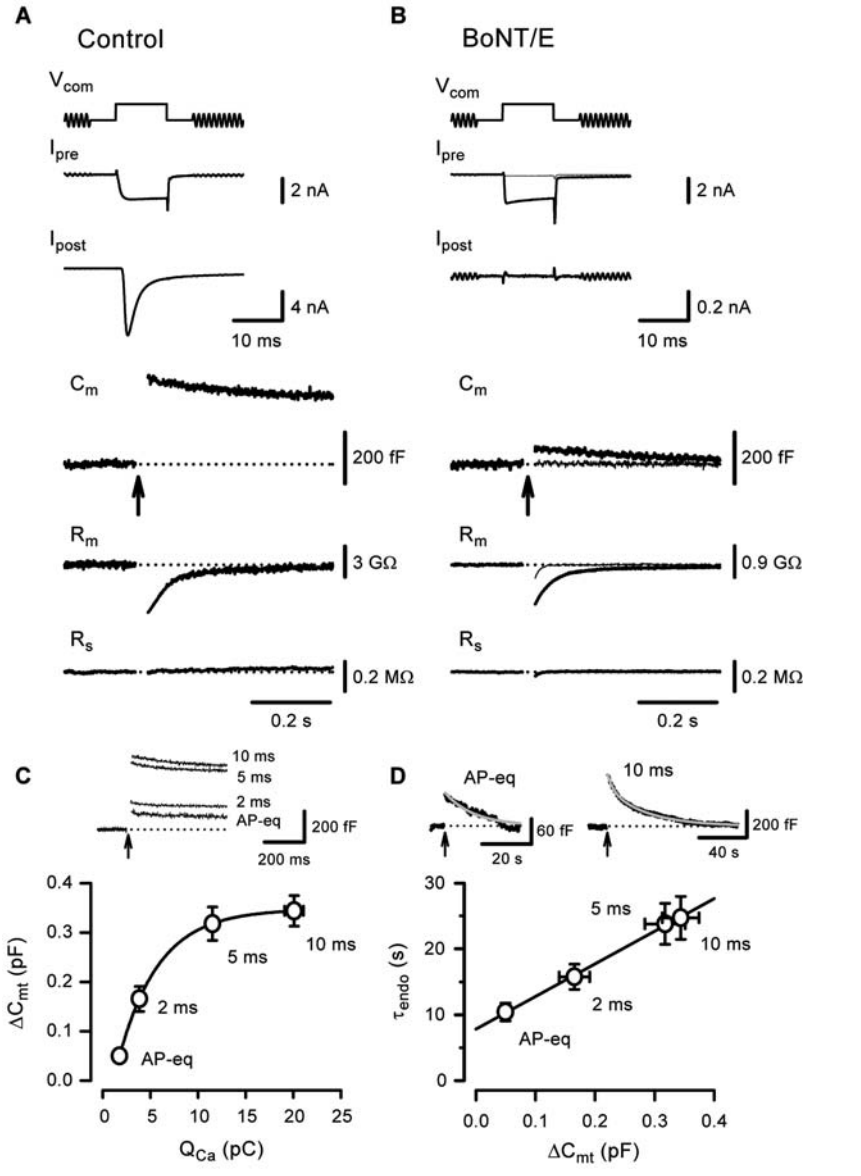


Fig. 1. Exo- and endocytosis of synaptic vesicles in simultaneous pre- and postsynaptic recordings at the calyx of Held. (**A**) Ca^{2+} currents (I_{pre}), EPSCs (I_{post}), and capacitance (C_m) change of presynaptic terminals induced by a depolarizing command pulse (V_{com}) (10 ms, from –80 to 0 mV). Membrane resistance (R_m) and series resistance (R_s) are simultaneously recorded. Baseline levels of C_m, R_m, and R_s were 24.0 pF, 5.5 giga-ohms (GΩ), and 14.6 mega-ohms (MΩ), respectively. (**B**) BoNT/E (200 nM) included in the presynaptic patch pipette abolished EPSCs (I_{post}) but spared changes in C_m and R_m, both of which were blocked by $CdCl_2$ (0.25 mM, thin traces, superimposed). Baseline levels of C_m, R_m, and R_s were 26.9 pF, 1.3 giga-ohms, and 10.3 mega-ohms, respectively. (**C**) The relationship between the amplitude of the release-related C_m change (ΔC_{mt}) and the charge of presynaptic Ca^{2+} currents (Q_{Ca}). Ca^{2+} currents were induced by 2- to 10-ms depolarizing pulses or AP-eq stimulation (*14*) (sample C_m traces in inset). Data points (n = 6 to 14) were fitted by a single exponential curve. (**D**) The relationship between the ΔC_{mt} decay time constant (τ_{endo}) and the ΔC_{mt} amplitude (n = 6 to 8, averaged C_m traces in inset). The ΔC_{mt} decay was fitted by a single (AP-eq) or double (10-ms) exponential curve (in gray, superimposed). Arrows indicate the onset of depolarizing pulse. Error bars show mean ± SEM.

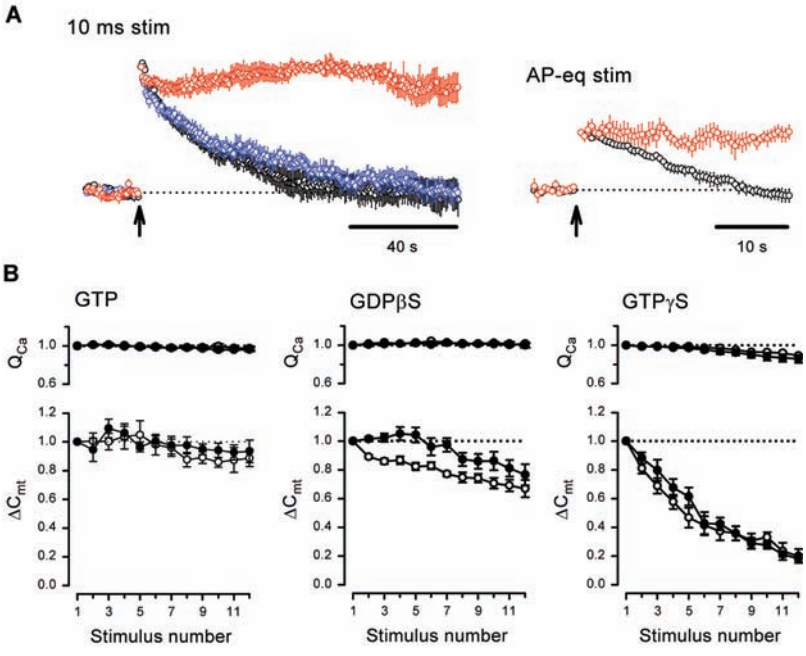


Fig. 2. Effects of guanine nucleotide analogs on vesicle endocytosis and recycling. (**A**) Averaged ΔC_{mt} traces evoked by 10-ms (left) or AP-eq (right) depolarizing pulse in the presence of GTP (0.3 mM, black, n = 5 to 6), GDPβS (3 mM, blue, n = 6), or GTPγS (0.2 mM, red, n = 4 to 5) in the presynaptic pipette. Averaged ΔC_{mt} traces are superimposed after normalizing the amplitude at 500 to 550 or 600 to 650 ms after depolarization. Arrows indicate the onset of depolarizing pulse. (**B**) ΔC_{mt} induced by a 10-ms pulse applied at 1-min intervals (open circles) or 2-min intervals (filled circles) in the presence of GTP, GDPβS, or GTPγS (n = 4 to 6) in the presynaptic pipette. Upper panels indicate presynaptic Ca^{2+} current charge (Q_{Ca}). Both ΔC_{mt} and Q_{Ca} are normalized to the initial value. Error bars show mean ± SEM.

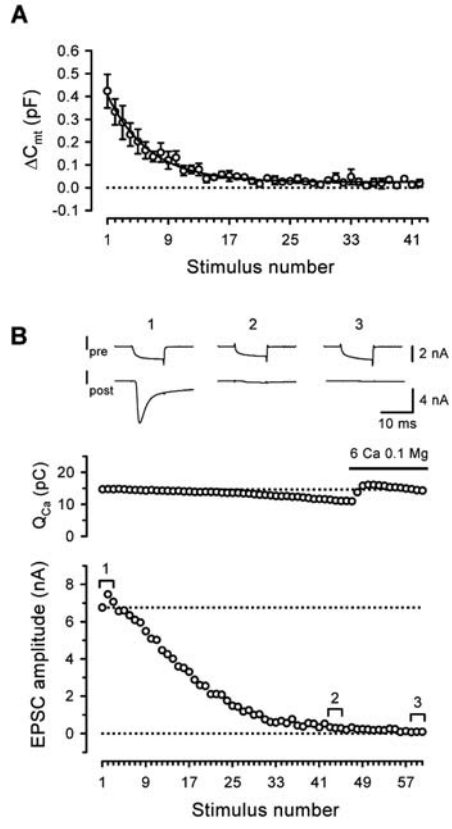


Fig. 4. Depletion of recycling vesicles by GTPγS in the presynaptic terminal. (**A**) Summary data of ΔC_{mt} (four calyces) evoked by the presynaptic Ca^{2+} currents every 1 min in the presence of GTPγS in the presynaptic pipette. (**B**) Top panel: EPSCs evoked every 1 min in the presence of GTPγS sampled from different epochs (1 to 3, bottom panel). After the 47th stimulation, extracellular Ca^{2+} concentration was raised to 6 mM (Mg^{2+}, 0.1 mM, bar), but no EPSC was evoked (3).

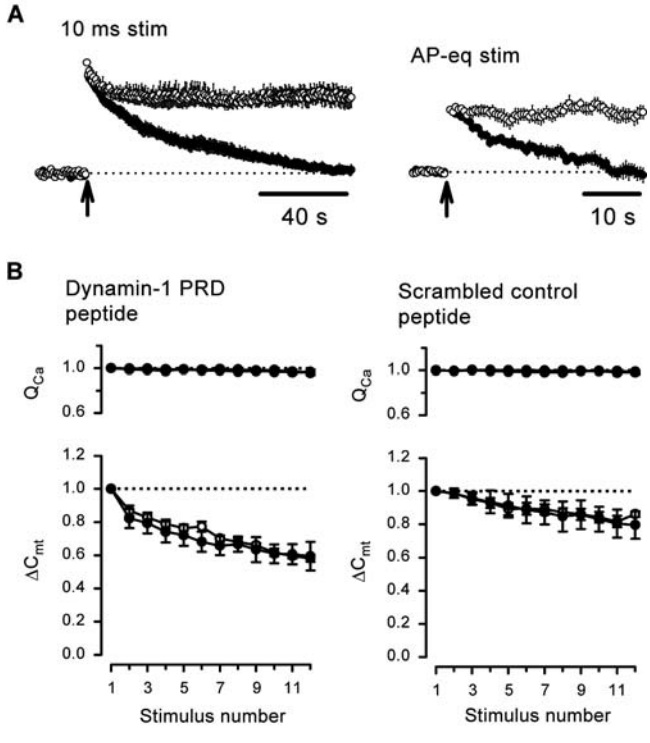


Fig. 3. Effects of dynamin-1 PRD peptide on vesicle endocytosis and recycling. (**A**) Averaged ΔC_{mt} evoked by 10-ms (left) or AP-eq (right) pulse in the presence of PRD peptide (open circles, n = 4 to 5) or scramble peptide (filled circles, n = 5) in the presynaptic pipette (both at 1 mM). Arrows indicate the onset of depolarizing pulse. (**B**) ΔC_{mt} induced by a 10-ms pulse applied at 1-min intervals (open circles) or 2-min intervals (filled circles) in the presence of PRD peptide (left, n = 4 to 5) or scramble peptide (n = 4). After establishing whole-cell recording, before stimulation, we allowed at least 13 min for peptides to diffuse into calyces. Error bars show mean ± SEM.

References and Notes

1. S. J. Royle, L. Lagnado, *J. Physiol.* **553**, 345 (2003).
2. T. C. Südhof, *Annu. Rev. Neurosci.* **27**, 509 (2004).
3. J. H. Koenig, K. Ikeda, *J. Neurosci.* **9**, 3844 (1989).
4. A. M. van der Bliek, E. M. Meyerowitz, *Nature* **351**, 411 (1991).
5. K. Takei, P. S. McPherson, S. L. Schmid, P. De Camilli, *Nature* **374**, 186 (1995).
6. R. Heidelberger, *J. Neurosci.* **21**, 6467 (2001).
7. A. M. Aravanis, J. L. Pyle, R. W. Tsien, *Nature* **423**, 643 (2003).
8. S. P. Gandhi, C. F. Stevens, *Nature* **423**, 607 (2003).
9. C. R. Artalejo, J. R. Henley, M. A. McNiven, H. C. Palfrey, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8328 (1995).
10. R. Fesce, F. Grohovaz, F. Valtorta, J. Meldolesi, *Trends Cell Biol.* **4**, 1 (1994).
11. J. Y. Sun, X. S. Wu, L. G. Wu, *Nature* **417**, 555 (2002).
12. O. Kjaerulff, P. Verstreken, H. J. Bellen, *Nature Cell Biol.* **4**, E245 (2002).
13. T. Takahashi, T. Hori, Y. Kajikawa, T. Tsujimoto, *Science* **289**, 460 (2000).
14. Materials and methods are available as supporting material on *Science* Online.
15. M. Wölfel, R. Schneggenburger, *J. Neurosci.* **23**, 7059 (2003).
16. T. Binz *et al.*, *J. Biol. Chem.* **269**, 1617 (1994).
17. Stable amplitude of this capacitance component during repeated stimulation with BoNT/E makes it unlikely that the capacitance change arose from predocked vesicles protected from BoNT/E.
18. H. Taschenberger, R. M. Leão, K. C. Rowland, G. A. Spirou, H. von Gersdorff, *Neuron* **36**, 1127 (2002).
19. L. G. Wu, W. J. Betz, *Neuron* **17**, 769 (1996).
20. S. Sankaranarayanan, T. A. Ryan, *Nature Cell Biol.* **2**, 197 (2000).
21. GTPγS attenuated presynaptic Ca^{2+} currents evoked by the 10-ms pulse (*13*) but did not reduce the first ΔC_{mt} (fig. S2A), suggesting that GTPγS stimulates exocytosis, as it does in secretory cells (*33*).
22. D. Grabs *et al.*, *J. Biol. Chem.* **272**, 13419 (1997).
23. O. Shupliakov *et al.*, *Science* **276**, 259 (1997).
24. G. J. Augustine *et al.*, *J. Physiol.* **520**, 33 (1999).
25. M. Krauss *et al.*, *J. Cell Biol.* **162**, 113 (2003).
26. At the beginning of stimulation, the EPSC rundown was unclear because of AMPA receptor saturation caused by a 10-ms depolarizing pulse (*30*, *34*).

27. Y. Kajikawa, N. Saitoh, T. Takahashi, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8054 (2001).
28. In this condition, where EPSCs could no longer be evoked, the rapid capacitance component remained observed (fig. S1C), confirming that it is unrelated to synaptic transmission.
29. R. P. J. de Lange, A. D. G. de Roos, J. G. G. Borst, *J. Neurosci.* **23**, 10164 (2003).
30. J. Y. Sun, L. G. Wu, *Neuron* **30**, 171 (2001).
31. T. Sakaba, E. Neher, *Neuron* **32**, 1119 (2001).
32. P. Holroyd, T. Lang, D. Wenzel, P. De Camilli, R. Jahn, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16806 (2002).
33. E. Neher, *J. Physiol.* **395**, 193 (1988).
34. T. Ishikawa, Y. Sahara, T. Takahashi, *Neuron* **34**, 613 (2002).
35. We thank S. Kozaki for kindly providing BoNT/E and M. Tachibana and T. Tsujimoto for comments on the manuscript. Supported by Grant-in-Aid for Specially Promoted Research from the Ministry of Education, Culture, Science, and Technology.

Supporting Online Material
www.sciencemag.org/cgi/content/full/307/5706/124/DC1
Materials and Methods
Figs. S1 and S2
References

3 August 2004; accepted 10 November 2004
10.1126/science.1103631

Spindle Multipolarity Is Prevented by Centrosomal Clustering

Nicholas J. Quintyne,[1] Janet E. Reing,[1] Diane R. Hoffelder,[1] Susanne M. Gollin,[2] William S. Saunders[1]*

Most tumor cells are characterized by increased genomic instability and chromosome segregational defects, often associated with hyperamplification of the centrosome and the formation of multipolar spindles. However, extra centrosomes do not always lead to multipolarity. Here, we describe a process of centrosomal clustering that prevented the formation of multipolar spindles in noncancer cells. Noncancer cells needed to overcome this clustering mechanism to allow multipolar spindles to form at a high frequency. The microtubule motor cytoplasmic dynein was a critical part of this coalescing machinery, and in some tumor cells overexpression of the spindle protein NuMA interfered with dynein localization, promoting multipolarity.

Hyperamplification of the centrosome has been observed in many tumor tissues and cell lines and is linked with both aneuploidy and tumorigenesis (*1*–*4*). The extent of genomic instability is generally correlated with the degree of centrosomal abnormalities (*3*–*5*). Furthermore, centrosome abnormalities are more severe in high-grade and recurrent tumors and in cell lines that show aggressive malignant phenotypes (*1*–*4*). In mitosis, supernumerary centrosomes can lead to an increase in spindle poles, and multipolar spindles are found in many cancer cell types (*6*).

Although centrosome amplification is clearly important for multipolar spindles, the presence of extra centrosomes does not always lead to multipolar spindle formation. Certain cell types can apparently suppress multipolarity and form a bipolar spindle during mitosis even though the centrosomes are amplified (*7*–*9*). Furthermore, some centrosomal defects induce centrosomal amplification without multipolarity (*10*). To suppress multipolar spindles, the cell could functionally silence the extra centrosomes, preventing them from forming a spindle pole and leaving only two centrosomes active. Alternatively, a cell could coalesce the extra centrosomes into only two functional spindle poles (*7*, *11*, *12*).

The spindle protein NuMA has been shown to be critical for spindle assembly (*13*–*17*), and the *NUMA1* gene maps to one of the most frequently amplified chromosomal segments in cancer cells (*18*, *19*). We examined two cell lines that exhibited relatively high amounts of NuMA expression (Fig. 1A) and that had ~20% multipolar spindles (Fig. 1, B and D). If the overexpression of NuMA is driving multipolarity, then reduction of NuMA by small interfering RNA (siRNA) should lead to a return of bipolar spindles. Three days after a single siRNA transfection (*20*), NuMA amounts were reduced (Fig. 1C and fig. S1), whereas amounts of the associated proteins dynein and dynactin were unchanged (fig. S2) (*16*). To examine only cells that received siRNA (~50%), we labeled the siRNA duplex with a fluorescein marker. In both the UPCI:SCC103 and UPCI:SCC078 oral cancer cell lines, transfection with the NuMA siRNA nearly eliminated multipolar spindles (Fig. 1, B and D). Similar results were observed for the SK-HEP-1 liver adenocarcinoma cell line (Fig. 1D). When the NuMA level was allowed to recover 10 days after a single siRNA treatment, the frequency of multipolarity returned to that of untreated cells. Thus, overexpression of NuMA perturbs the ability of these cells to coalesce supernumerary centrosomes into a single pole. NuMA provides a cohesive force in maintaining spindle microtubules around a single centrosome (*13*, *16*), but these results suggest a previously unknown role for elevated amounts of NuMA as an inhibitor of centrosome coalescence in cells with supernumerary centrosomes.

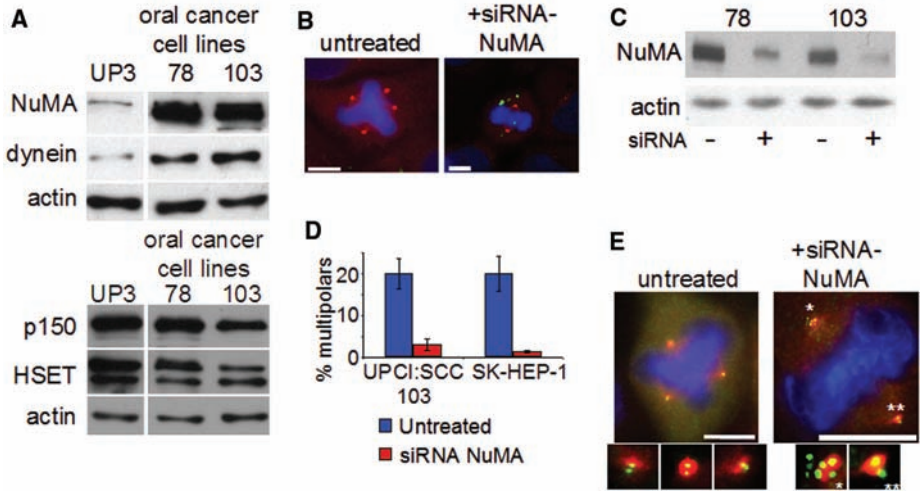


Fig. 1. Reduction of NuMA decreases spindle multipolarity. (**A**) Immunoblots of whole-cell extracts from the indicated cell lines were probed with antibodies to NuMA, dynein intermediate chain, the dynactin subunit p150^{Glued}, and HSET. Only NuMA and dynein showed an increase in expression in cancer cells when compared to normal oral keratinocytes (UP3) (*24*). (**B**) UPCI:SCC103 cells were stained with antibodies to centrosomal γ-tubulin (red), fluorescein-labeled siRNA (green), and the DNA dye 4′,6′-diamidino-2-phenylindole (DAPI) (blue) before and after siRNA transfection. (**C**) NuMA protein reduction after transfection. (**D**) Decrease in multipolar spindles in metaphase cells after siRNA to NuMA. In each case, the decrease in multipolar spindles was matched by an increase in bipolar spindles. (**E**) UPCI:SCC103 cells stained with antibodies to centrin-2 (green) and γ-tubulin (red) and with DAPI (blue). Magnified views of spindle poles are shown at bottom at 3× magnification. Asterisks match the magnified images with the larger image. Bars indicate 10 μm in all figures.

[1]Department of Biological Sciences, [2]Department of Human Genetics and the University of Pittsburgh Cancer Institute, University of Pittsburgh, 4249 Fifth Avenue, Pittsburgh, PA 15260, USA.

*To whom correspondence should be addressed. E-mail: wsaund@pitt.edu

Does NuMA act directly on the spindle or exert its effect on other spindle proteins? Dynein (*13*, *16*), its associated activator dynactin (*21*), and the kinesin motor HSET (*22*) all localized along the spindle microtubules in noncancer HEK293 cells (Fig. 2) (*22*, *23*). Whereas HSET and dynactin localized normally in the oral cancer cells, spindle dynein immunoreactivity was absent or sharply reduced (Fig. 2), although dynein intermediate chain expression was elevated in these cells (Fig. 1A) (*24*). Punctate dynein staining was still visible in the cancer cells, probably representing nonspindle vesicle-associated motor molecules. Reduction of NuMA did not interfere with dynein localization at the centrosomes of interphase cancer cells. A survey of various cancer cell types showed a similar loss of spindle dynein, but not dynactin, immunolabeling from nearly all cultured cancer cell lines tested (table S1).

Because NuMA overexpression led to multipolar spindle formation and dynein plays a critical role in spindle formation and maintenance, we tested whether the overexpression of NuMA was the cause of the change in spindle-associated dynein. When NuMA amounts were reduced in the UPCI:SCC103 oral cancer cell line (or UPCI:SCC078), dynein was visible on nearly all of the spindles (Fig. 3A, image e). More than 90% of the spindles in UPCI:SCC103 cells lacked dynein staining before treatment, whereas after siRNA treatment >80% of the transfected cells had visible dynein staining (Fig. 4A). Thus, overexpression of NuMA induced both a change in spindle dynein and multipolarity in these cancer cell lines. Similar results were observed for the SK-HEP-1 liver adenocarcinoma cell line (table S1).

However, spindle dynein delocalization in other cancer cell lines is not dependent on NuMA. In the UPCI:SCC070 cells, which do not have *NUMA1* amplification (*19*), reduction of NuMA had no effect on spindle multipolarity and did not restore dynein staining (Fig. 3A, image f). Similar observations were made in other tumor cell lines (table S1). Although a reduction of spindle dynein staining seems to be a common feature of many cancer cell lines, only some cell types achieve this change by overexpression of NuMA.

Could a change in spindle dynein account for the failure of centrosome clustering in cancer cells? We first checked for a correlation between the frequency of multipolar spindles and centrosomal amplification or dynein localization in various cancer cell lines. Of the tested lines, only those with centrosomal amplification and no dynein immunolabeling on the mitotic spindle were able to induce multipolar spindles in >10% of the metaphase cells (table S2). However, other unknown variables could also account for the differences between these different cell lines.

To confirm that these two factors were critical for spindle multipolarity, we determined whether inducing extra centrosomes and/or inhibiting dynein function would result in multipolarity in noncancer cells. We treated HEK293 cells with the microtubule inhibitor Colcemid (Irvine Scientific, Santa Ana, CA) for 28 to 36 hours, and the frequency of cells with extra centrosomes increased from <5% to ~80% (Fig. 3B, images a and a', and fig. S3). The increase in spindle multipolarity was limited to between ~8% and ~20% of the metaphase cells (Fig. 4, A and B; this would still be relatively high for the untreated cancer cells). Thus, centrosome amplification alone leads to only a limited increase in multipolarity in this nontumor cell line. To inhibit dynein, we transfected the HEK293 cells with either plasmids expressing NuMA (*14*) to mimic the overexpression seen in the oral cancer cells or a plasmid expressing the dynein-binding fragment of p150^{Glued}, CC1 (*25*). Overexpression of NuMA was able to displace dynein from the spindle of ~50% of the HEK293 cells, reproducing the change we saw in cancer cells with *NUMA1* amplification, but by itself did not increase the frequency of multipolar spindles (Figs. 3B, images b and b' and 4, B and C) (*21*). Similarly, expression of CC1 reduced dynein on the spindle as expected but only marginally elevated spindle multipolarity (Fig. 4, B and C). Thus, inhibition of dynein or amplification of centrosomes alone was sufficient to induce only a modest increase in multipolar spindles. However, when cells were treated with both Colcemid and overexpression of either NuMA or CC1, multipolar spindle frequency increased to ~60 to 70% of metaphase cells (Figs. 3B, images c and c' and 4, B and C).

We also tried two other methods of amplifying centrosomes in conjunction with inhibition of dynein. HEK293 cells were transfected with a plasmid expressing the centrosomal kinase hMps1 (*26*). Centrosomal amplification rose from 6.1% to 36% of the cells, but multipolarity only increased when cells were cotransfected with plasmids expressing NuMA or CC1 (fig. S4, A and B). Similarly, the preexisting centrosomal coalescence of N1E-115 cells (*8*, *9*) was eliminated by NuMA or CC1 overexpression, and spindle multipolarity jumped from ~5% to ~80%. Another human oral cancer cell line (UPCI:SCC114) was found to possess similar, but less pronounced, centrosomal

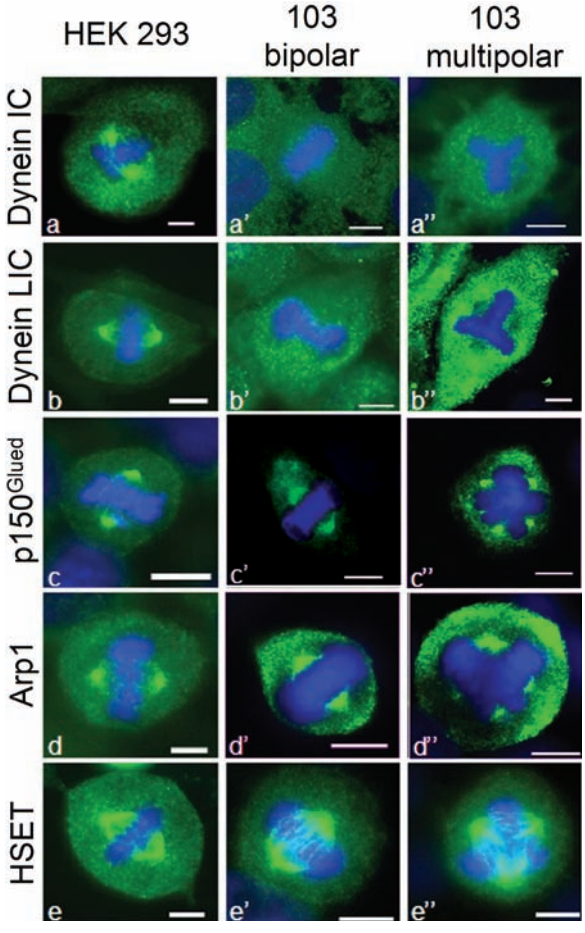


Fig. 2. Dynein is depleted from the spindle in oral cancer cells. HEK293 (images a to e) and UPCI:SCC103 (images a' to e' and a'' to e'') cells were stained with antibodies to dynein, the p150^{Glued} or Arp 1 subunits of dynactin, or HSET (green) and with DAPI (blue). Spindles in UPCI:SCC103 cells lacked visible dynein but were positive for dynactin and HSET. Antibodies to two different subunits of the dynein macromolecular complex gave similar results, indicating that the loss of immunoreactivity was not likely to be caused by epitope masking.

clustering like that of N1E-115. Unlike the other human cancer lines we tested that did not show centrosomal clustering, expression of either NuMA or CC1 alone in UPCI:SCC114 markedly elevated spindle multipolarity in the absence of Colcemid exposure or hMps1 overexpression (fig. S4, C to E).

It appears that spindle multipolarity arises via two distinct steps, an increase in centrosome number and an inhibition of centrosomal coalescence (*7*, *11*, *12*). Centrosome hyperamplification has been described in numerous tumor types (*1*–*4*). However, a second change leading to a loss of centrosome coalescence was required to manifest the multipolar phenotype. Thus, clustering may be an important mechanism for preserving genomic stability in noncancer cells. Overcoming centrosomal clustering appears to involve a change in spindle dynein, either a reduction of dynein amounts or a change to a more diffuse position within the spindle such that the strong fluorescent signal is not seen. It is unlikely that dynein is completely inhibited in the spindle of cancer cells, because we did not see the splaying of spindle poles observed after injection of antibodies to dynein (*21*) and we observed NuMA labeling on the spindle, which requires transport by dynein (*27*). Apparently, enough dynein activity remains to prevent these phenotypes. However, these results show that dynein plays an important role in maintaining the coalescing mechanism to prevent multipolar spindles.

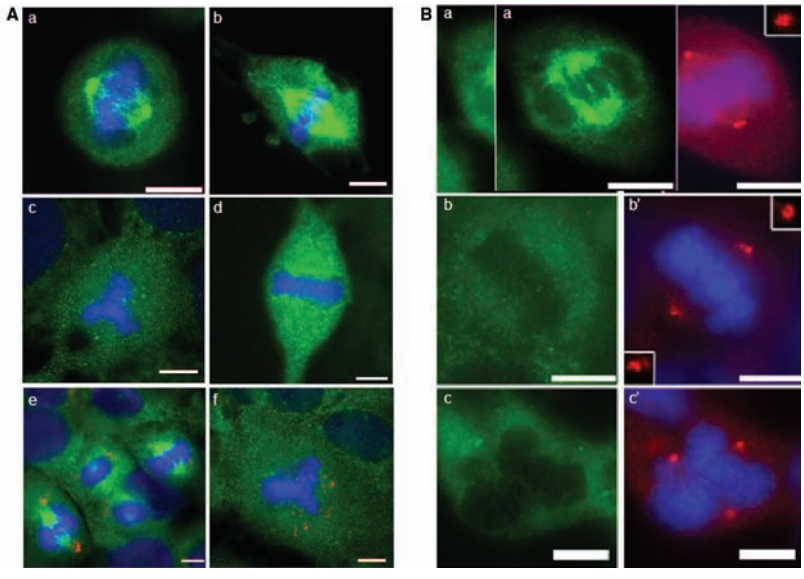


Fig. 3. Dynein is restored to spindles after siRNA-mediated knockdown of NuMA. (**A**) Cells stained with antibodies to dynein intermediate chain (green) and DAPI (blue). Normal human oral keratinocytes (UP3) (image a) and skin fibroblasts (image b) showed the expected spindle-associated dynein. Untransfected metaphase UPCI:SCC103 showed little or no detectable dynein on multipolar (image c) or bipolar spindles (image d). Dynein returned to spindles in UPCI:SCC103 cultures transfected with NuMA siRNA (red, image e) but not for UPCI:SCC070 (image f). (**B**) Inhibition of dynein or NuMA overexpression stimulated formation of multipolar spindles only in cells with amplified centrosomes. HEK293 cells were treated with Colcemid (a and a'), or overexpression of NuMA (images b and b'), or both (images c and c'). Antibodies used in (A) images and (B) images a to c were anti-dynein LIC (green) and in (B) images a' to c' were anti-γ-tubulin (red). DAPI, blue. Insets, 3× magnifications.

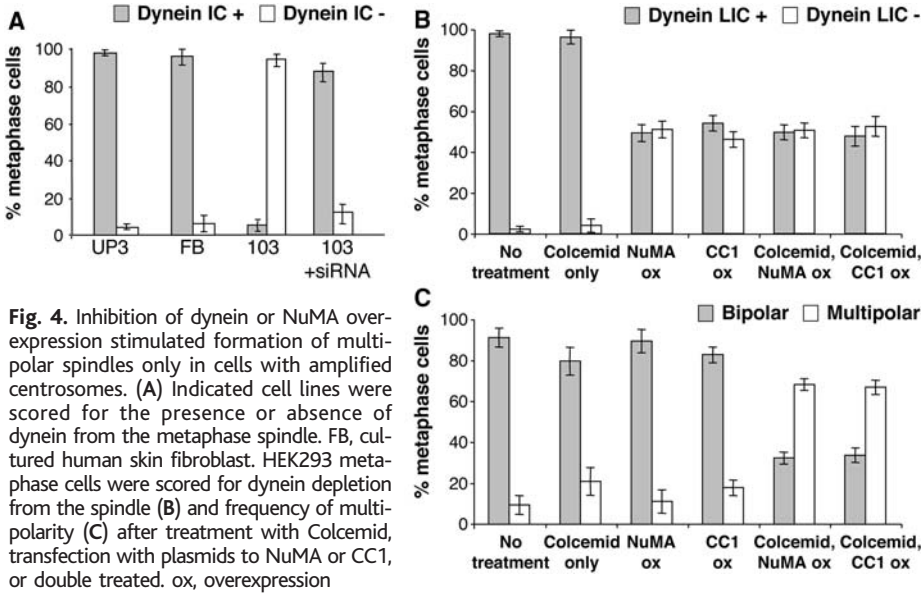


Fig. 4. Inhibition of dynein or NuMA overexpression stimulated formation of multipolar spindles only in cells with amplified centrosomes. (**A**) Indicated cell lines were scored for the presence or absence of dynein from the metaphase spindle. FB, cultured human skin fibroblast. HEK293 metaphase cells were scored for dynein depletion from the spindle (**B**) and frequency of multipolarity (**C**) after treatment with Colcemid, transfection with plasmids to NuMA or CC1, or double treated. ox, overexpression

References and Notes

1. P. E. Carroll *et al.*, *Oncogene* **18**, 1935 (1999).
2. A. B. D'Assoro, W. L. Lingle, J. L. Salisbury, *Oncogene* **21**, 6146 (2002).
3. N. Sato *et al.*, *Cancer Genet. Cytogenet.* **126**, 13 (2001).
4. G. A. Pihan *et al.*, *Cancer Res.* **61**, 2212 (2001).
5. A. B. D'Assoro *et al.*, *Breast Cancer Res. Treat.* **75**, 25 (2002).
6. W. S. Saunders *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 303 (2000).
7. B. R. Brinkley, *Trends Cell Biol.* **11**, 18 (2001).
8. D. Ring, R. Hubble, M. Kirschner, *J. Cell Biol.* **94**, 549 (1982).
9. G. A. Sharp, K. Weber, M. Osborn, *Eur. J. Cell Biol.* **29**, 97 (1982).
10. S. Duensing *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10002 (2000).
11. G. Sluder, J. J. Nordberg, *Curr. Opin. Cell Biol.* **16**, 49 (2004).
12. E. A. Nigg, *Nat. Rev. Cancer* **2**, 815 (2002).
13. T. Gaglio *et al.*, *J. Cell Biol.* **135**, 399 (1996).
14. D. A. Compton, D. W. Cleveland, *J. Cell Biol.* **120**, 947 (1993).
15. D. A. Compton, C. Luo, *J. Cell Sci.* **108**, 621 (1995).
16. A. Merdes, K. Ramyar, J. D. Vechio, D. W. Cleveland, *Cell* **87**, 447 (1996).
17. T. Gaglio, A. Saredi, D. A. Compton, *J. Cell Biol.* **131**, 693 (1995).
18. S. M. Gollin, *Head Neck* **23**, 238 (2001).
19. X. Huang, S. M. Gollin, S. Raja, T. E. Godfrey, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11369 (2002).
20. S. M. Elbashir *et al.*, *Nature* **411**, 494 (2001).
21. A. Merdes, R. Heald, K. Samejima, W. C. Earnshaw, D. W. Cleveland, *J. Cell Biol.* **149**, 851 (2000).
22. V. Mountain *et al.*, *J. Cell Biol.* **147**, 351 (1999).
23. T. Yoshida, A. Ito, K. Izutsu, *Cell Struct. Funct.* **10**, 245 (1985).
24. R. K. Vadlamudi *et al.*, *Cancer Cell* **5**, 575 (2004).
25. N. J. Quintyne *et al.*, *J. Cell Biol.* **147**, 321 (1999).
26. H. A. Fisk, C. P. Mattison, M. Winey, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14875 (2003).
27. T. Gaglio, M. A. Dionne, D. A. Compton, *J. Cell Biol.* **138**, 1055 (1997).

28. The authors thank D. Compton for antibodies to NuMA and HSET; A. Merdes for the pGW-*NUMA1* plasmid; T. Schroer for the CC1 plasmid and DLIC and Arp1 antibodies; M. Winey for the hMps1 plasmid; Z. Yu for culturing the UP3 cells; and X. Huang, J. Brodsky, and V. Nanda for helpful discussions. The work was supported by NIH grant DE016086 and American Cancer Society grant RSG-96-016-06 to W.S.S. and the Oral Cavity Cancer Center suppported by NIH grant P60DE13059.

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5706/127/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 and S2

7 September 2004; accepted 3 November 2004
10.1126/science.1104905

The Centromeric Protein Sgo1 Is Required to Sense Lack of Tension on Mitotic Chromosomes

Vahan B. Indjeian, Bodo M. Stern,* Andrew W. Murray†

Chromosome alignment on the mitotic spindle is monitored by the spindle checkpoint. We identify Sgo1, a protein involved in meiotic chromosome cohesion, as a spindle checkpoint component. Budding yeast cells with mutations in *SGO1* respond normally to microtubule depolymerization but not to lack of tension at the kinetochore, and they have difficulty attaching sister chromatids to opposite poles of the spindle. Sgo1 is thus required for sensing tension between sister chromatids during mitosis, and its degradation when they separate may prevent cell cycle arrest and chromosome loss in anaphase, a time when sister chromatids are no longer under tension.

Errors in chromosome segregation lead to disease and death. To prevent such errors, the protein complex cohesin (*1*) holds replicated chromosomes together, and this linkage is not broken until every pair of sister chromatids is bi-oriented on the mitotic spindle, with the two sisters attached to microtubules that emanate from opposite poles of the spindle (*2*). Bi-orientation generates tension on the chromosomes because the links between the sister chromatids resist the pulling forces of the spindle. Chromosome orientation is monitored by the spindle checkpoint, which detects unattached kinetochores (*3*) or the lack of tension between sister chromatids (*4*, *5*). Either lesion inhibits the signal that induces chromosome segregation. Several components of the spindle checkpoint (Mad2, Mad3, and Bub3) form a complex that prevents entry into anaphase by binding to Cdc20, an essential activator of the anaphase-promoting complex (APC) (*6*). The APC triggers the destruction of securin (Pds1), the protein that inhibits separase (Esp1), which is the protease that triggers sister separation by cleaving cohesin (*1*) (Fig. 1A). When the spindle checkpoint inhibits the APC, Pds1 is stable, separase is inhibited, and the sisters remain unseparated.

Little is known about how cells sense the absence of tension between a pair of sister chromatids and send this information to the spindle checkpoint. Ipl1, the budding yeast member of the Aurora family of mitotic protein kinases (*7*), is involved in this process (*8*) and is also needed to detach microtubules from kinetochores that are not under tension (*9*–*11*). The only role of Ipl1 may be to induce chromosomes to detach from microtubules, thus creating naked kinetochores, which inhibit the APC by recruiting other checkpoint components. Alternatively, the lack of tension at the kinetochore may signal directly to the checkpoint without requiring microtubule detachment (*8*, *12*, *13*).

To identify components of the tension-sensing machinery, we looked for mutants that ignored chromosomes that were not under tension. The screen used a budding yeast (*Saccharomyces cerevisiae*) strain harboring linear minichromosomes (LMCs) that segregate poorly and activate the spindle checkpoint (*14*). In wild-type budding yeast, lack of tension delays APC activation, but cells eventually enter anaphase (*5*, *8*). This delay becomes a lethal arrest in cells that contain *CDC28-VF*, a mutation in Cdc28 (the budding yeast homolog of the cyclin dependent kinase Cdk1) that reduces APC activity (*15*). This arrest is dependent on the spindle checkpoint and can be overcome by expressing *CDC20-127*, a dominant Cdc20 allele that is refractory to Mad2 inhibition and therefore overrides the checkpoint (*16*) (Fig. 1A). We mutagenized the *CDC28-VF* strain carrying the LMCs and selected mutants that grow in the absence of the checkpoint-resistant Cdc20 and therefore bypass the minichromosome-induced mitotic arrest (Fig. 1A). Two of the mutants lie in *SGO1*, the fungal homolog of *Drosophila MEI-S332*, a gene whose product protects centromeric cohesion during the first meiotic division (*17*–*20*). Each allele carries a mutation in the most conserved region of the protein ($Thr^{379} \rightarrow$ Ile in *sgo1-100* and $Pro^{390} \rightarrow$ His in *sgo1-700*); the *sgo1-700* allele also has a second substitution ($Asp^{519} \rightarrow$ Asn) in a region that is strongly conserved among fungi.

Sgo1 is cell cycle regulated. Its expression increased at the G_1-S transition and decreased during mitosis in a pattern similar to that of Pds1 (Fig. 1B). Like Pds1, Sgo1 was stabilized in the presence of benomyl and nocodazole, drugs that destabilize microtubules (Fig. 1C). Sgo1 could be a substrate of either the APC or separase. Because Sgo1 was still degraded in cells that expressed a nondegradable version of Pds1 that inhibits separase but has no effect on the APC (*21*) (Fig. 1D), Sgo1 is likely a direct target of the APC. Four potential APC recognition motifs (D-boxes) in Sgo1 further support this possibility.

Our selection was designed to identify components that signal and/or sense tension at the kinetochores. To examine the response to chromosomes that lack tension, we placed *sgo1* mutants in cells that also express Mcd1 (also known as Scc1), a component of cohesin (*1*), under the glucose-repressible *GAL1* promoter. In the presence of glucose, Mcd1 is not expressed and the sister chromatids are not linked. The positioning of the sisterless chromosomes on the spindle indicates that they attach to the spindle (*5*), although we cannot prove that every kinetochore is attached to a microtubule at all times. In the absence of Mcd1, reduced tension at the kinetochores activates the spindle checkpoint and delays APC activation (*5*, *8*), as indicated by stabilization of the APC target Pds1 (Fig. 2A). The *sgo1* mutants, however, failed to delay APC activation, hence they are unable to sense or respond to lack of tension.

We also examined cells lacking Sgo1 (*sgo1*Δ). In the W303 strain background, *sgo1*Δ cells showed poor viability and acquired suppressor mutations (*22*). Reduced viability was largely overcome by slowing down DNA replication, an approach previously used to reduce the sensitivity of spindle checkpoint mutants to benomyl (*23*). The exact timing of Pds1 degradation was somewhat variable in *sgo1*Δ cells, but within an experiment the timing of destruction was similar regardless of whether Mcd1 was expressed (Fig. 2A); this result implies that the *sgo1*Δ cells cannot sense the lack of tension. The *sgo1* point mutants also failed to delay APC activation in cells lacking Cdc6 expression (*22*). Cdc6 is required for DNA replication, and in its absence chromosomes enter mitosis without the sisters needed to generate tension at the kinetochore (*5*, *8*).

The *sgo1* mutants shared properties with other spindle checkpoint mutants such as *mad2*Δ. The *sgo1* mutants showed chromosome loss rates (*24*) approximately equal to those of *mad2*Δ cells, five times the rate for wild-type cells, and one-third the rates for *bub1*Δ and *bub3*Δ strains (table S1) (*25*). Like *mad2*Δ, the growth of the *sgo1* mutants was sensitive to benomyl (Fig. 2B). To examine their response to microtubule depolymerization, we arrested cells in G_1 by

Department of Molecular and Cellular Biology, Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA.

*Present address: Cell Press, 1100 Massachusetts Avenue, Cambridge, MA 02138, USA.
†To whom correspondence should be addressed. E-mail: amurray@mcb.harvard.edu

exposure to mating pheromone (α-factor) and then released them into medium containing benomyl and nocodazole. The viability of *sgo1* mutants dropped markedly after microtubule depolymerization (Fig. 2C) even though the *sgo1* mutants, like wild-type cells, arrested with high levels of Pds1 for at least 4 hours (Fig. 2D). In contrast, *mad2*Δ cells failed to detect the damaged spindle and degraded Pds1 about 2 hours after their release from G_1 arrest (Fig. 2D). Therefore, the *sgo1* mutants, unlike *mad2*Δ, are still able to arrest in mitosis in response to unattached kinetochores. Although we cannot exclude the possibility that weak defects in the spindle checkpoint allow cells to respond to unattached kinetochores, but not to those that are not under tension, the observation that *sgo1* mutants that have different effects on viability are all inactivate to tension sensing without altering the ability to detect unattached kinetochores argues against this possibility.

The behavior of *sgo1* mutants seems paradoxical; they die after microtubule depolymerization even though they arrest. Because Sgo1 protects centromeric cohesion during meiosis (*17*–*19*), one explanation is that the *sgo1* mutants have mitotic cohesion defects. We introduced the *sgo1* alleles into strains that have chromosome IV labeled with 256 tandem repeats of the Lac operator inserted at the *TRP1* locus (12 kb from the centromere). This array is seen as a fluorescent dot when cells express a green fluorescent protein–Lac repressor fusion protein (GFP-LacI) (*26*). As controls, we used wild-type, *mad2*Δ, and *GAL1-MCD1* strains in the same background. The *mad2*Δ cells lack all known aspects of the spindle checkpoint, and *GAL1-MCD1* cells lack cohesin when grown in glucose. The strains were released from G_1 into medium that contained glucose, benomyl, and nocodazole. In wild-type cells, the absence of kinetochore-microtubule attachment activates the spindle checkpoint, stabilizing cohesion. As a result, sister chromatids stay together and the GFP-LacI dots corresponding to the two sisters of chromosome IV cannot be resolved, appearing as a single dot under the fluorescent microscope. Roughly half the cells lacking Mad2 or cohesin had two green dots, showing that they had prematurely separated their sisters (Fig. 3A). The *sgo1* mutants behaved much more like wild-type cells and predominantly arrested as large-budded cells with a single visible dot.

These results show that *sgo1* mutants do not have a major mitotic cohesion defect. Their rapid death in the absence of microtubules could be explained in two ways. One is that the *sgo1* mutants have a minor cohesion defect that is small enough to differentiate them from cells lacking cohesin, yet ensures that at least one chromosome missegregates in most benomyl- and nocodazole-treated cells. The other is that cohesion is normal, but *sgo1* mutants make errors in chromosome alignment as the spindle reforms after benomyl and nocodazole have been removed. We distinguished these hypotheses by delaying sister chromatid segregation until well after benomyl and nocodazole had been removed, a treatment that has no effect on chromosomes that have already separated but allows more time for linked pairs of sisters to align correctly on the spindle.

We compared *sgo1-100*, wild-type, and *mad2*Δ cells, all containing the GFP-marked chromosome IV. Two versions of each strain were made, differing only in the presence of an extra copy of the *MPS1* gene driven by the *GAL1* promoter. Mps1 is part of the spindle checkpoint, and its overexpression causes a spindle checkpoint and APC-dependent metaphase arrest (*27*). The APC and Mps1 mutually oppose each other because the APC induces the destruction of Mps1 (*28*). Hence, the metaphase arrest induced by overexpressing Mps1 can be rapidly reversed by adding glucose. All the strains were released from G_1 into medium that contained galactose, benomyl, and nocodazole. Regardless of whether they overexpressed Mps1, wild-type

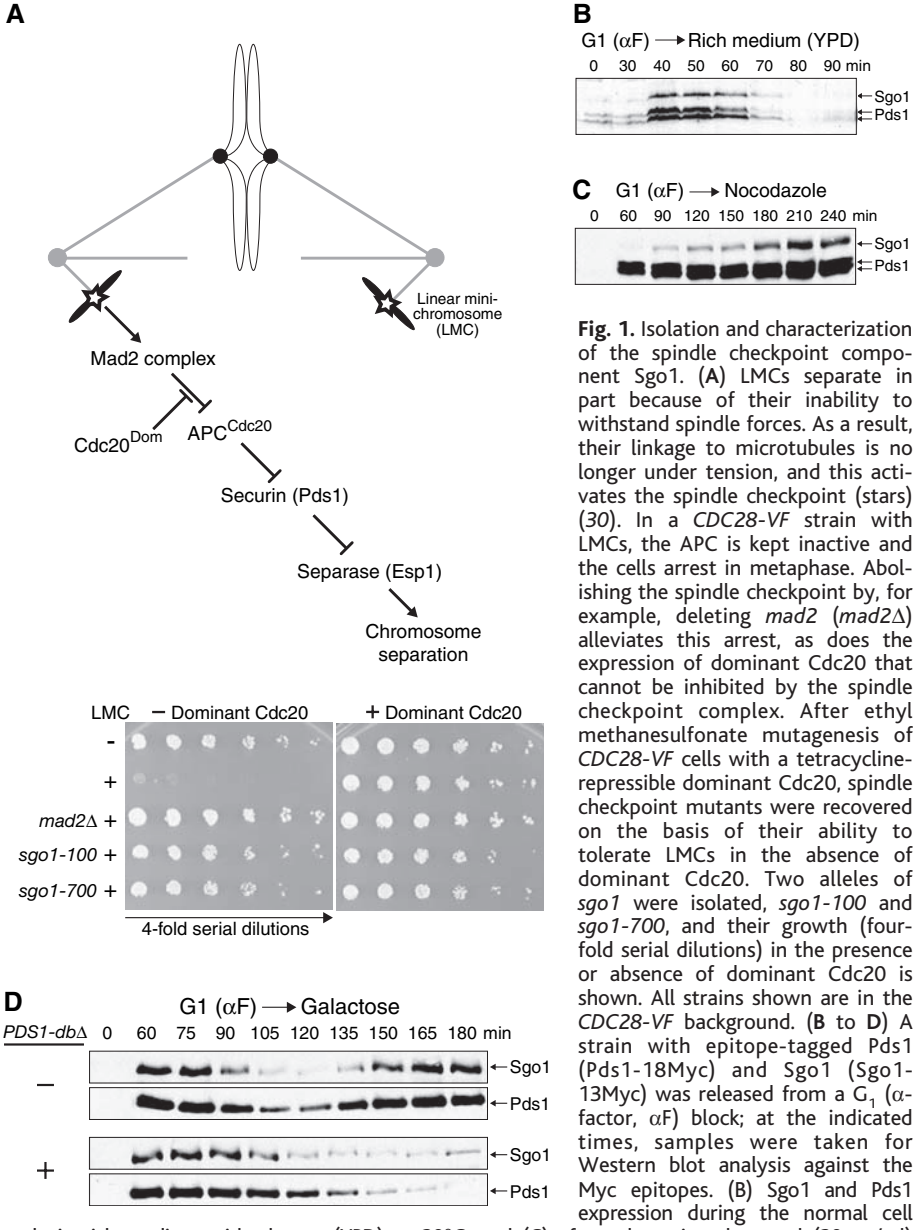


Fig. 1. Isolation and characterization of the spindle checkpoint component Sgo1. (**A**) LMCs separate in part because of their inability to withstand spindle forces. As a result, their linkage to microtubules is no longer under tension, and this activates the spindle checkpoint (stars) (*30*). In a *CDC28-VF* strain with LMCs, the APC is kept inactive and the cells arrest in metaphase. Abolishing the spindle checkpoint by, for example, deleting *mad2* (*mad2*Δ) alleviates this arrest, as does the expression of dominant Cdc20 that cannot be inhibited by the spindle checkpoint complex. After ethyl methanesulfonate mutagenesis of *CDC28-VF* cells with a tetracycline-repressible dominant Cdc20, spindle checkpoint mutants were recovered on the basis of their ability to tolerate LMCs in the absence of dominant Cdc20. Two alleles of *sgo1* were isolated, *sgo1-100* and *sgo1-700*, and their growth (four-fold serial dilutions) in the presence or absence of dominant Cdc20 is shown. All strains shown are in the *CDC28-VF* background. (**B** to **D**) A strain with epitope-tagged Pds1 (Pds1-18Myc) and Sgo1 (Sgo1-13Myc) was released from a G_1 (α-factor, αF) block; at the indicated times, samples were taken for Western blot analysis against the Myc epitopes. (B) Sgo1 and Pds1 expression during the normal cell cycle in rich medium with glucose (YPD) at 30°C, and (C) after release into benomyl (30 μg/ml) and nocodazole (30 μg/ml) at 23°C; (D) Sgo1 and Pds1 expression after release from G_1 to galactose at 30°C in the presence or absence of nondegradable, destruction box–deleted Pds1 (Pds1-dbΔ) under control of the *GAL1* promoter. In both cases the APC is activated, as evidenced by the destruction of the endogenous Pds1-18Myc. However, in the presence of Pds1-dbΔ the strains remained arrested as large budded cells, whereas in its absence the cells rebudded after about 120 min (*22*).

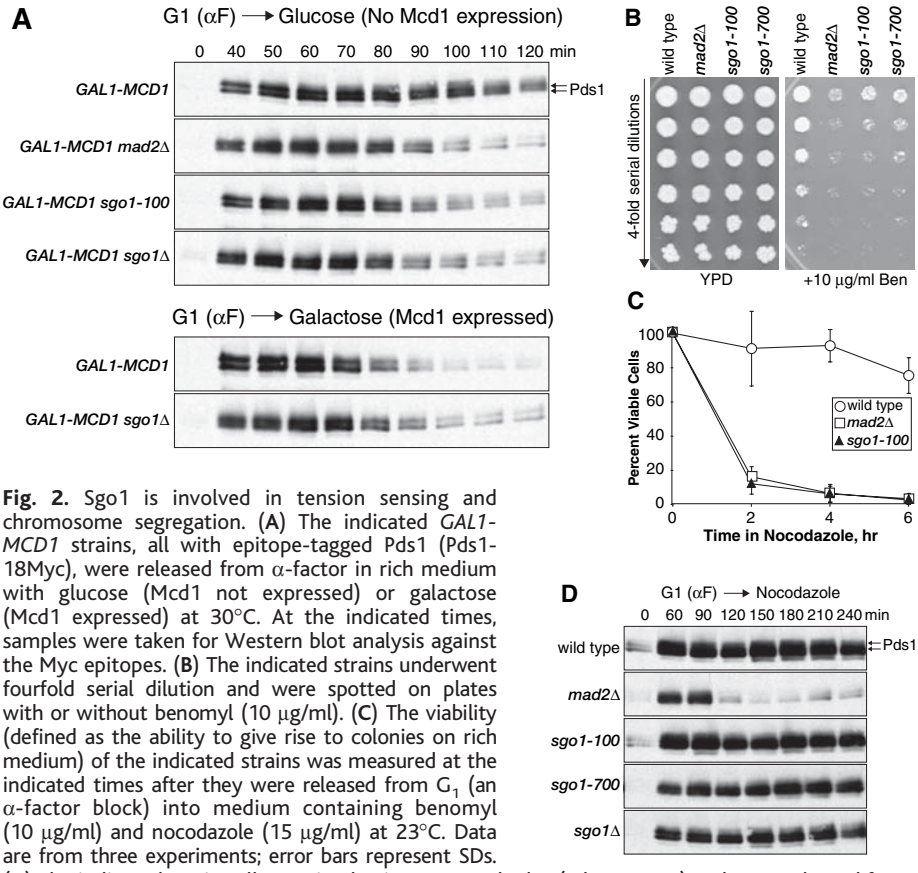


Fig. 2. Sgo1 is involved in tension sensing and chromosome segregation. (**A**) The indicated *GAL1-MCD1* strains, all with epitope-tagged Pds1 (Pds1-18Myc), were released from α-factor in rich medium with glucose (Mcd1 not expressed) or galactose (Mcd1 expressed) at 30°C. At the indicated times, samples were taken for Western blot analysis against the Myc epitopes. (**B**) The indicated strains underwent fourfold serial dilution and were spotted on plates with or without benomyl (10 μg/ml). (**C**) The viability (defined as the ability to give rise to colonies on rich medium) of the indicated strains was measured at the indicated times after they were released from G_1 (an α-factor block) into medium containing benomyl (10 μg/ml) and nocodazole (15 μg/ml) at 23°C. Data are from three experiments; error bars represent SDs. (**D**) The indicated strains all contained epitope-tagged Pds1 (Pds1-18Myc) and were released from α-factor in media containing benomyl (30 μg/ml) and nocodazole (30 μg/ml) at 23°C. Samples were taken at the indicated times for Western blot analysis against the Myc epitopes.

and *sgo1-100* cells arrested in metaphase, whereas *mad2*Δ cells did not and separated their sisters prematurely (*22*). The cells were then transferred to medium that contained galactose but lacked benomyl and nocodazole for 1.5 hours to allow the spindle to reassemble well before chromosome segregation (Fig. 3B). In this experiment, overexpression of Mps1 increased the viability of *sgo1-100* cells but had no effect on *mad2*Δ cells whose sisters had already separated. To confirm that rescue depended on delaying anaphase until the spindle had reformed, we transferred *sgo1-100* cells that had been treated with galactose, benomyl, and nocodazole into medium containing glucose, benomyl, and nocodazole for 0.5 hours, allowing Mps1 levels to fall before washing out the microtubule-destabilizing drugs. These cells died, indicating that Mps1 overexpression was only effective when it continued after the benomyl and nocodazole had been removed (Fig. 3B). We conclude that although *sgo1* mutants arrest normally when they lack a spindle, they die because they make but cannot detect errors in chromosome alignment as the spindle reforms.

We monitored chromosome segregation as cells recovered from microtubule depolymerization. Cells that had overexpressed Mps1 were released from their final mitotic block into α-factor, which prevents their escape from G_1, and numbers of GFP dots (a proxy for the number of copies of chromosome IV) per cell were counted. Bi-

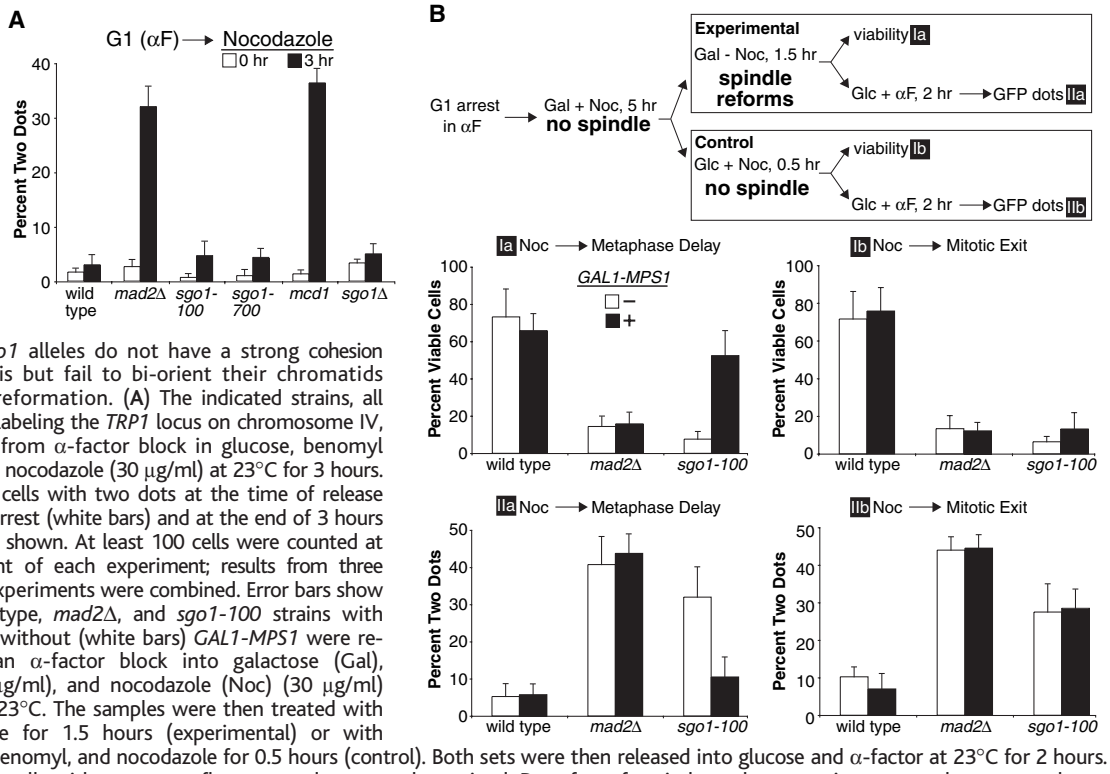


Fig. 3. The *sgo1* alleles do not have a strong cohesion defect in mitosis but fail to bi-orient their chromatids after spindle reformation. (**A**) The indicated strains, all with GFP dots labeling the *TRP1* locus on chromosome IV, were released from α-factor block in glucose, benomyl (30 μg/ml), and nocodazole (30 μg/ml) at 23°C for 3 hours. Percentages of cells with two dots at the time of release from α-factor arrest (white bars) and at the end of 3 hours (black bars) are shown. At least 100 cells were counted at each time point of each experiment; results from three independent experiments were combined. Error bars show SDs. (**B**) Wild-type, *mad2*Δ, and *sgo1-100* strains with (black bars) or without (white bars) *GAL1-MPS1* were released from an α-factor block into galactose (Gal), benomyl (30 μg/ml), and nocodazole (Noc) (30 μg/ml) for 5 hours at 23°C. The samples were then treated with galactose alone for 1.5 hours (experimental) or with glucose (Glc), benomyl, and nocodazole for 0.5 hours (control). Both sets were then released into glucose and α-factor at 23°C for 2 hours. Percentages of viable cells and cells with two green fluorescent dots were determined. Data from four independent experiments are shown; error bars show SDs.

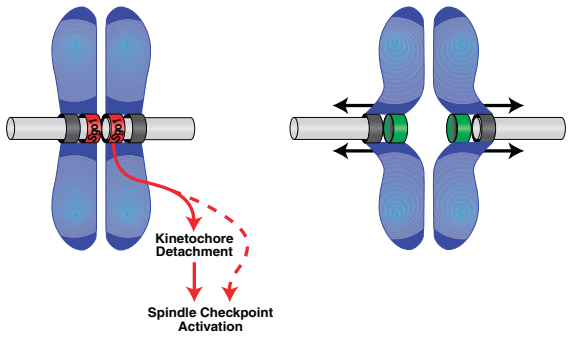


Fig. 4. A model of Sgo1 function in sensing and/or signaling tension and promoting bi-orientation of mitotic chromosomes (blue). Spindle microtubules (light gray) regulate Sgo1 activity, which switches between an active (red) and inactive (green) state. In the absence of tension, Sgo1 is in contact with spindle microtubules and sends signals to promote bi-orientation and delay progression into anaphase. The dashed arrow represents uncertainty about whether the lack of tension can inhibit APC activity without generating unattached kinetochores. A separate mechanism (dark gray) may monitor attachment of kinetochores to spindle microtubules.

orientation of a chromosome leads to each daughter cell inheriting one sister chromatid and thus having a single GFP dot. Failure to bi-orient the sister chromatids leaves one daughter with two GFP dots and its sister with none. As expected from the rescue of viability, allowing Mps1 overexpression as the spindle reformed substantially reduced the fraction of *sgo1-100* cells that had two dots in the following G_1 (Fig. 3B). These results argue that *sgo1-100* cells cannot properly align their sister chromatids on the spindle but do successfully hold them together.

We detected two new defects in *sgo1* mutants: an inability to arrest the cell cycle in cells whose chromosomes were not under tension, and a defect in chromosome segregation that could be rescued by delaying the onset of anaphase. Vertebrate homologs of Sgo1 contain a strong microtubule-binding domain (*29*), and Sgo1 is found at the kinetochore (*18*). Both observations argue that Sgo1 is directly involved in interactions between chromosomes and microtubules, and raise the possibility that it is the tension sensor. For example, Sgo1 could be located in a region of the kinetochore that could only be reached by microtubules that were not under tension (Fig. 4). As long as it is bound to a microtubule, Sgo1 would send a signal that destabilizes the attachment of the kinetochore to its microtubule and might also send a signal directly to the spindle checkpoint. Sister centromeres separate from each other as anaphase begins. Once chromosomes have left their sisters, arresting the cell cycle would be pointless and the destabilization of microtubule-kinetochore attachment could have disastrous consequences for chromosome segregation. The degradation of Sgo1 as anaphase begins and the resulting ablation of the tension-signaling pathway may explain why sister chromatids remain stably attached to the spindle once they reach its poles.

References and Notes

1. K. Nasmyth, *Science* **297**, 559 (2002).
2. A. Musacchio, K. G. Hardwick, *Nature Rev. Mol. Cell Biol.* **3**, 731 (2002).
3. C. L. Rieder, R. W. Cole, A. Khodjakov, G. Sluder, *J. Cell Biol.* **130**, 941 (1995).
4. X. Li, R. B. Nicklas, *Nature* **373**, 630 (1995).
5. B. M. Stern, A. W. Murray, *Curr. Biol.* **11**, 1462 (2001).
6. H. Yu, *Curr. Opin. Cell Biol.* **14**, 706 (2002).
7. M. Carmena, W. C. Earnshaw, *Nature Rev. Mol. Cell Biol.* **4**, 842 (2003).
8. S. Biggins, A. W. Murray, *Genes Dev.* **15**, 3118 (2001).
9. I. M. Cheeseman *et al.*, *Cell* **111**, 163 (2002).
10. S. Biggins *et al.*, *Genes Dev.* **13**, 532 (1999).
11. T. U. Tanaka *et al.*, *Cell* **108**, 317 (2002).
12. S. Hauf *et al.*, *J. Cell Biol.* **161**, 281 (2003).
13. C. Ditchfield *et al.*, *J. Cell Biol.* **161**, 267 (2003).
14. W. A. Wells, A. W. Murray, *J. Cell Biol.* **133**, 75 (1996).
15. A. D. Rudner, K. G. Hardwick, A. W. Murray, *J. Cell Biol.* **149**, 1361 (2000).
16. L. H. Hwang *et al.*, *Science* **279**, 1041 (1998).
17. V. L. Katis, M. Galova, K. P. Rabitsch, J. Gregan, K. Nasmyth, *Curr. Biol.* **14**, 560 (2004).
18. T. S. Kitajima, S. A. Kawashima, Y. Watanabe, *Nature* **427**, 510 (2004).
19. A. L. Marston, W. H. Tham, H. Shah, A. Amon, *Science* **303**, 1367 (2004).
20. A. W. Kerrebrock, W. Y. Miyazaki, D. Birnby, T. L. Orr-Weaver, *Genetics* **130**, 827 (1992).
21. R. L. Tinker-Kulberg, D. O. Morgan, *Genes Dev.* **13**, 1936 (1999).
22. V. B. Indjeian, A. W. Murray, unpublished data.
23. R. Li, A. W. Murray, *Cell* **66**, 519 (1991).
24. P. Hieter, C. Mann, M. Snyder, R. W. Davis, *Cell* **40**, 381 (1985).
25. See supporting data on *Science* Online.
26. A. F. Straight, A. S. Belmont, C. C. Robinett, A. W. Murray, *Curr. Biol.* **6**, 1599 (1996).
27. K. G. Hardwick, E. Weiss, F. C. Luca, M. Winey, A. W. Murray, *Science* **273**, 953 (1996).
28. W. Palframan, B. M. Stern, A. W. Murray, unpublished data.
29. A. Salic, J. C. Waters, T. J. Mitchison, *Cell* **118**, 567 (2004).
30. B. M. Stern, K. Liem, A. W. Murray, unpublished data.
31. We thank S. Davis for invaluable technical assistance with the selection; all members of the Murray lab for helpful comments and suggestions; A. Amon and T. Mitchison for sharing data before publication; S. Biggins and A. Amon for helpful comments; and R. Hellmiss-Peralta for advice on graphics. Supported by NIH grant GM043987 (A.W.M.) and by a Leukemia and Lymphoma Society fellowship (B.M.S.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/307/5706/130/DC1
Materials and Methods
Tables S1 and S2

14 June 2004; accepted 6 November 2004
10.1126/science.1101366

Kinase Profiling System

The SignalScout Kinase Profiling System enables users to study phosphorylation by protein kinases in a multiplex format, while also providing the means to build their own kinase assays. The SignalScout Kinase Substrate Ladder Kit provides turnkey multiplex kinase assays for convenient and specific detection of the activity of four kinases in one experiment. The collection of individual kinase substrates allows users to customize their own assays. There is also a line of antibodies against protein kinases for protein immunoblotting applications.

Stratagene For information 858-535-5400 www.stratagene.com

Toxic Effects Assays

The ATPLite is an adenosine triphosphate (ATP) monitoring system based on firefly luciferase. This luminescence assay is an alternative to colorimetric, fluorometric, and radioisotopic assays for the quantitative evaluation of proliferation and cytotoxicity of cultured mammalian cells. ATP monitoring can be used to assess the cytocidal, cytostatic, and proliferative effects of a wide range of drugs, biological response modifiers, and biological compounds. The major advantages of the system are high sensitivity, excellent linearity, simplicity, fast results, and the lack of cell harvesting or separation steps.

PerkinElmer For information 800-551-2121 www.perkinelmer.com

Contact Printing Kits

These Contact Printing Kits allow researchers to transfer a chemical or biological agent from a low energy polydimethylsiloxane (PDMS) surface to a high-energy gold or glass printing surface. In addition, gold-coated printing surfaces can be functionalized. Each kit includes the user's choice of three PDMS stamps, printing surfaces, and thiols to functionalize printing surfaces. These kits will enable scientists to print biological molecules onto a high-energy surface, micro- and nano-pattern chemistries onto a surface, and print molecules onto a self-assembled monolayer (SAM) or pattern SAMs.

Platypus Technologies For information 866-296-4455 www.platypustech.com

T Cell Receptor Antibodies

In the ongoing expansion of a line that now includes more than 20 T cell receptor (TCR) antibodies, anti-human TCR Vα2, Vγ9, and Vδ2 monoclonal antibodies are now available in both unlabeled and fluorescein-labeled forms. Applications include flow cytometry, immunoprecipitation, and immunohistochemistry. TCRs are important in research into autoimmune disease, T cell lymphoma, tumor suppression and rejection, multiple sclerosis, rheumatic heart disease, heart transplant rejection, leukemia, and other disorders.

Pierce Biotechnology For information 800-487-4885 www.endogen.com

Advanced Systems Biology Tools

Two systems biology companies are integrating and co-marketing tools for functional analysis of high-throughput experimental data. Ariadne Genomics' MedScan TextMiner automated text processing pipeline will be integrated with MetaCore, GeneGo's flagship analytical platform that includes proprietary, manually curated databases of human biology and chemistry. MetaCore features maps, pathways, and networks for almost all known human proteins; parsers for importing data from all major microarray platforms as well as proteomics and metabolomics data; seven proprietary network building algorithms; and flexible visualization and data exchange tools. MedScan TextMiner makes use of a proprietary natural language processing engine to automatically process Medline abstracts, articles, and other scientific text in TXT, PDF, HTML, and other formats. It extracts functional associations among proteins, cell processes, and small molecules, and recognizes types of regulatory mechanisms involved and the effects of regulation.

Ariadne Genomics and GeneGo For information 847-644-1557 www.adriadnegenomics.com

Leak-Free Tubing

Pre-assembled Omnifit tubing sets enable instrument manufacturers to simplify product assemblies, eliminate assembly errors, ensure leak-free connections, reduce assembly time, and condense purchasing and inventory management tasks. Tubing sets are cut into the appropriate lengths and assembled with the new distinctive, permanently attached, color-coded Ferrulok fitting system on one or both ends. Combinations of multiple and single tubing sets can be prepared to plumb complete instruments.

Bio-Chem Valve For information 973-263-3001 www.bio-chemvalve.com

Method Development HPLC Systems

Updated versions of a line of automated method development high performance liquid chromatography (HPLC) systems incorporates more user-defined features and control options. The systems feature enhanced functionality and simplified operation to enable users of various experience levels to accelerate their pharmaceutical development process. With just a few mouse clicks, users can maintain complete control from injection to detection to report generation and e-mail notification. Analytical and preparative systems are available. The analytical system functions as a true method development tool that facilitates characterizations of microgram to milligram quantities of lead candidates. The innovative Stacked Plot Report overlays chromatograms to find the perfect set of conditions from method development runs on multiple columns. An identical interface in the preparative system allows the two systems to work side by side, first establishing analysis conditions, then scaling up to enable the isolation of target compounds by a variety of detectors.

Shimadzu For information 800-477-1227 www.shimadzu.com

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and government organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier by visiting www.science.labvelocity.com on the Web, where you can request that the information be sent to you by e-mail, fax, mail, or telephone.

Published by AAAS